

REFERENCE LECTURE NOTE



Principles and Practices of

Fundamentals and Clinical Epidemiology

Salaudin Myia
Binod Regmi

Principles and Practices of **Fundamentals and Clinical Epidemiology**

(Reference Lecture Note)

Salaudin Myia

MPH, MA

Head of Department

Hope International College

Binod Regmi

MPH, MA, LLB

Head of Department

National Academy for Medical Sciences

Developed as reference note for students of public health by compilation of contents from different published and unpublished resource. This lecture note acknowledge to all these resources. The contents in this lecture note are not own thoughts of writers whereas the lecture note reserved the copy right

Table of Contents

CHAPTER 1	6
Concept of Epidemiology	6
1.1. Introduction	6
1.2. Historical development of epidemiology	6
1.3. Definitions of Epidemiology	7
1.4. Components of epidemiology	9
1.5. Aims of epidemiology	10
1.6. Scope of epidemiology	10
1.7. Principles/Tenets of epidemiology	11
1.8. Uses of epidemiology	11
1.9. Contributions of Epidemiology in Medical Field	12
1.10. Spectrum of health and disease	12
1.11. Determinant of Health and Disease	13
1.12. Basic Epidemiological Measurements	17
CHAPTER 2	30
Epidemiological Methodology/Studies (Study Design)	30
2.1 Classification of epidemiological study designs	30
2.2 Descriptive study	32
2.3 Analytical study designs	37
2.3.1 Cross sectional studies	37
2.3.2 Ecological study	38
2.3.3 Longitudinal studies	38
2.3.3.1 Case control study	39
2.3.3.2 Cohort Study	44
2.4 Experimental Studies	48
2.4.1 True Experimental studies	48
2.4.2 Quasi experimental studies	51
2.5 Strategies of epidemiology	52
CHAPTER 3	55
Infectious Disease Epidemiology	55
3.1. Terminology Used In Epidemiology	55
3.2. Mode of Disease Transmission	58
3.3. Epidemiological triad	61

3.4. Disease Prevention and Control	63
CHAPTER 4	65
Clinical Epidemiology	65
4.1. Introduction of Clinical Epidemiology	65
4.2. Screening	69
Advantages	74
Disadvantages	75
Community-Based Screening for Chronic Kidney Disease by BPKIHS	76
4.3. Diagnostic test function and confounding-	76
4.4. Yield	83
4.5. Precision	83
4.6. Assumption of independence	84
What is the Assumption of Independence?	84
What happens if violate the Assumption of Independence?	86
How do Avoid Violating the Assumption?	86
4.7. Confounding	86
4.8. Normality and Abnormality	88
4.9. ROC Curve	91
4.10. Description of performance of diagnostic and calculation	94
4.11. Surveillance	101
4.12. Strategies of epidemiology	104
4.13. Lead time of diseases mechanism	105
4.14. Concept of risk, uses of risk, studies of risk and risk factors	106
4.15. Latency period	108
4.16. Concept of prognosis of disease and prognostic factors	109
4.17. Emerging and re-emerging infectious diseases and threat to clinical and public health	110
CHAPTER 5	113
Genetic Epidemiology	113
Modern concept of gene:	114
Gene Types	115
1. Single gene inheritance	115
2. Multifactorial inheritance	116
3. Chromosome abnormalities	116

4. Mitochondrial inheritance	116
How are genetic problems diagnosed?	117
What are teratogenic problems?	117
CHAPTER 6	118
Social Epidemiology	118
CHAPTER 7	126
Field Epidemiology	126
7.1. Introduction	126
7.2. Field Techniques (Steps of proposal development)	127
References	141

CHAPTER 1

Concept of Epidemiology

1.1. Introduction

Epidemiological speculations and research findings often blaze across media headlines and arouse public hysteria. Some questions raised include what is more dangerous: vaccination against smallpox or infection with disease itself. The Assumption regarding the use of concept of epidemiology has been traced with Greek word. The word epidemiology is derived from Greek word epidemic means

Epi: among

Demos: people

Logos: study, knowledge, science

The term epidemiology began with Adam and Eve, both trying to investigate the quality of forbidden fruits which dating back to third century B.C. There was widespread and heavy toll of communicable diseases and its gloomy picture of its consequences were concern of ancient epidemics. There were many deaths, low survival rate with high proportion of population having disabilities. Consideration of the existing condition, the initial concentration of epidemiology was directed to infectious diseases and the epidemic conditions and their management. The subject matter of study remained on communicable diseases.

Epidemiology is the study of occurrence, distribution and determinants of health problems and diseases among human population and use of this knowledge for prevention, control and treatment of diseases. Thus, the primary unit of epidemiology is group of people and population can be defined in terms of geographical areas, specific age and attribute groups. So, epidemiology is a multidisciplinary subject

involving the ideas of many sectors such as biostatistics, research, clinical medicine, veterinary sciences, psychology and other social sciences.

1.2. Historical development of epidemiology

Early causal explanation for the epidemic of disease was the beliefs on supernatural forces (wrath of god, Breakdown of religious belief, consequences of past bad works) with no empirical evidences other than supernatural forces. Hippocrates who is popularly known as father of medicine expressed the term epidemiology over 2000 years ago. He attempted to explain the occurrences of diseases form rational cause instead of supernatural force. In his essay entitled "On air, water and place" suggested that environmental and host factors might have influences in the development of disease. It was concluded that environmental factors were the factors in disease causation.

John Graunt, who studied the mortality statistics, published mortality data in 1662. He is the first person to quantify the pattern of birth, death and disease occurrences, infant mortality, urban/rural and seasonal variation of these events. He utilized the mortality counts for the description of health status of population. Then, an anesthesiologist John Snow

conducted several classical investigations in London around mid 1800s; probably, twenty years before the development of microscope. He performed epidemiological investigation of cholera outbreak (1848-1854 AD) to discover the cause of disease and to prevent recurrence under natural conditions. He sketched the spot map and pointed out the sources of drinking water to illustrate the outbreak and the area affected by cholera epidemic. The source of infection for the cholera was suspected to be polluted water. He made an explanation on the basis of characteristics in terms of time, place and person. Then he tried to test it by applying more comprehensive studies. John Snow proved the association of polluted water with cholera through natural experimentation.

William Farr during 1800s built upon the Graunt's works by systematically collecting and analyzing the mortality statistics in Britain. His study was concentrated on collection, analysis and evaluation of data on Vital statistics and disease classification.

During mid and late 1800s many other European and United States sphere countries began to apply the epidemiological methods for the study of chronic diseases and others non infectious diseases among the human population. In USA, Winslow and Sedgwick taught epidemiology in the early 1920s, although the subject was not given departmental status. Then after in 1927, W.H. Frost became the first professor of epidemiology in USA and then after epidemiology has been evolved as a subject matter of study in medical colleges.

Study by Doll and Hill (1950s), linking to smoking and cancer and study of cardiovascular diseases among the residents of Framingham, Massachusetts are two pioneering researches since Second World War. Finally the epidemiological methods and principles were applied for the eradication of smallpox during 1960s and 1970s.

Epidemiology has been evolved rapidly during past three decades, with its many ramifications which cover the study of diseases and other health and health related events by identifying etiology / risk factors of chronic diseases, formulating and evaluating treatment modalities.

1.3. Definitions of Epidemiology

A) Historical Definition

- Epidemiology is the science of epidemic and endemic disease - Stedman's Medical Dictionary
- The science of infective disease, their prime causes propagation and prevention - Stallybrass 1931
- Epidemiology is the branch of medical science, which treats of epidemics - Parkin 1873.
- Epidemiology is the science of mass phenomena of infections disease. - Frost (1927)

B) Modern Definition

- Epidemiology is the study of distribution and determinants of disease frequency in man - Mac Mohan (1960).
- The latest definition of epidemiology, which encompasses its modern application,

emphasizes that epidemiology is the study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to control health problems. - John M Last (1988).

Many definitions proposed but the following captures the principles and public health spirit of epidemiology.

“Epidemiology is the quantitative study of the distribution and determinants of health-related states or events in specified populations, and the application of this study to the control of health problems.” (John M Last 2006).

Short Explanation of terms used in the definition of epidemiology is stated below.

- Quantitative
It implies measurement of particular health states and of hazards (exposures, behaviours etc.), and the counting of individuals.
- Study
Epidemiology is a scientific discipline, sometimes called —the basic science of public health.|| It has, at its foundation, sound methods of scientific inquiry. It includes surveillance, observation, hypothesis testing, and analysis by time, place and person and other in age/sex distribution.
- Distribution
Epidemiology is concerned with the frequency and pattern of health events in a population. Frequency includes not only the number of such events in a population, but also the rate or risk of disease in the population. The rate (number of events divided by size of the population) is critical to epidemiologists because it allows valid comparisons across different populations. Pattern refers to the occurrence of health-related events by time, place, and personal characteristics.
- Determinants
Determinants include physical, biological, social, cultural and behavioral influence on health.
- Health-related states
It includes the conditions of health, disease, disability, discomfort dissatisfaction, health needs and demands, health activities, health care facilities, health care utilization and life style etc.
- Health related events
Health related events include the vital events like birth, death, marriage, divorce, migrations, calamities such as flood, famine, drought, earth quake and other accidents fall injuries injury by animals etc.
- Specified populations
Mean population in as defined geographical area with specific characteristics
- Control of health problems
To make explicit aim that epidemiology prevents, promotes and protects /restores the life of individuals /groups

Modern epidemiology

Modern epidemiology evolved with the study by Doll and Hill. It is the study of illness and can be defined as a “systematic body of epidemiologic principles by which to design and

judge (epidemiologic) studies have begun to form only in the past decades” - Rothman - 1986

Modern epidemiology deals with, in addition to the infectious diseases, chronic diseases and others health related states that is not a medical condition but can give rise to health consequences. The concept has started with the search of numbers of communicable diseases and expanding its horizon towards genetic epidemiology, social epidemiology, pharmaco-epidemiology, immuno-epidemiology, Geographical epidemiology, clinical epidemiology, field epidemiology etc.

1.4. Components of epidemiology

Although there has been no absolute definition and components of epidemiology that all epidemiologists subscribe, there is a consensus that following three components are common to all. These are

1. Disease frequency
2. Disease Distribution
3. Determinants of Disease

A brief explanation of each component is as follows.

1. Disease frequency

Inherent to the definition of epidemiology, it is the measurement of frequency of disease, disability and deaths and summarizing this information in terms of rate, ratio and Proportion. Thus, the rate and ratio are the basic measurements of frequency of events. This information is required for the comparison of such findings in similar conditions. Equally, epidemiology is important to the study of health related states /events such as blood pressure, height, weight, health need health care utilization etc. In this concept, the epidemiology is the quantitative science and Biostatistics is the basic tool of epidemiology. Example:

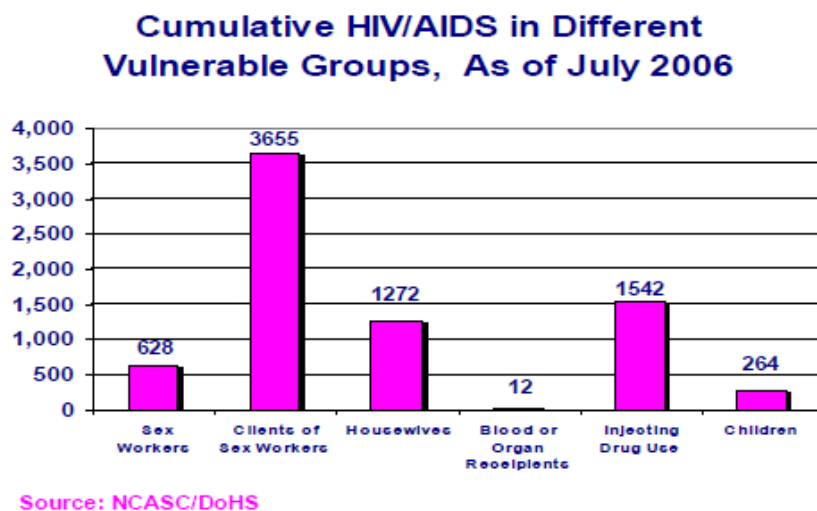


Fig: Frequency of HIV infection among different groups of people

2. Distribution of Disease

It is obvious that disease is not uniformly distributed among all. There are variations

in many attributes (age, sex, socioeconomic status occupation etc). An important function of epidemiology is to study these patterns among various groups, subgroups of population by time, place and person i.e. it studies whether the disease has been increased or decreased over the period of time. Thus, it is known as descriptive epidemiology. The distribution is divided into three categories namely time, place and person distribution. Time distribution is most important aspect of epidemiological distribution. it states –when does a disease occur?|| answer may be expressed in terms of time factor e.g. annual attack rate, incidence rate of month/year etc. Place distribution can be expressed in terms of regional, local or global coverage by disease/events and person distribution usually refers to the personnel characteristics of individuals /groups which are attributable to health related events or states Example: lung cancer is more prevalent in male while females suffer more frequently from breast cancer. Similarly, Negros are resistant to Falciparum Malaria because of absence of an enzyme G6PD.

3. Determinants of disease

An important function of epidemiology is to test etiological hypothesis and to identify the underlying cause/s of disease/event. In order to do so, epidemiological principles must be followed. This is the core of epidemiology which is known as analytical epidemiology.

1.5. Aims of epidemiology

According to the international epidemiological association, epidemiology has three main aims

- To describe the distribution and size of problem in the specified human population
- To identify the etiological factors in the pathogenesis of disease
- To provide the data essential for planning, implementation and evaluation of health services for the prevention, control and treatment of disease to the setting of priorities among those services.

Thus, the ultimate aim of epidemiology is to eliminate or reduce the health problem and to promote health and well being of society.

1.6. Scope of epidemiology

The accurate definition and determinants of health are not clear yet. Definition and dimensions of health are illusive. So, there is further broad fields of epidemiology that are still remaining unexplored. The study of epidemiology is not only concerned with death, illness and disability but also extended to other health and health related aspects too. It looks forward for the positive state of health the scope of epidemiology can be traced in terms of its applicable points.

1. As planning tools
 - Assess the problems in population
 - Trend of the problems
 - Compare the problems
 - Design interventions
2. As management tool

- Evaluate programs
 - Monitor the progress
 - Output assessment
 - Impact assessment
- Surveillance of disease
- Investigation of outbreak
- 3. As advocacy tool
 - Policy advocacy
 - Intervention choices

1.7. Principles/Tenets of epidemiology

Fundamental points to be followed in epidemiology are as follows

1. The target of the study in epidemiology is human population
2. All findings in epidemiology must be related to defined population
3. Conclusions are based on comparison
4. There are six WH questions namely
 - When did it happen? -----Time
 - Where did it happen----- Place
 - Who are affected ----- Person
 - Others are what, how and why

1.8. Uses of epidemiology

Epidemiology is most useful in the public health in assisting its mission, aims and operations of protecting the health of the populations and sub-populations. The most uses of epidemiology are given below.

- i. Natural history of disease
Epidemiology studies the trends of a disease and predicts the disease trends. The results of epidemiological studies are useful in planning for health programs and services.
- ii. Community diagnosis
Epidemiology determines the diseases, conditions, injuries, disorders, disabilities, and defects causing illness, health problems or deaths in a community or regions.
- iii. Risk factors of the diseases
Epidemiology looks the risk factors of the diseases, problems, behaviors of the affected groups. Groups are studied by doing risk factors assessments and health appraisal approaches. E.g. health risk, appraisal, health screening, medical exams, disease assessment.
- iv. Assessment, evaluation and research
Epidemiology also assesses the impact of the health programs/services to meet the goal and objectives (meeting the problems of the populations). It helps on effectiveness, efficacy, quality, quantity, access, and availability of services to treat, control, or prevent diseases. It also studies the injury, disability or death in the population or sub group of population.
- v. Completing the clinical picture
Identification or diagnostic processes to establish that conditions exist or that a person has a specific disease or ill- health condition. It also studies the cause -effect

relationships. E.g. sore throat can cause the rheumatic disease.

vi. Identification of syndromes

Epidemiology helps to establish and set criteria to define syndromes. E.g. sudden death in infants, fetal alcohol etc.

vii. Determining the cause and effect, source of disease.

Epidemiological findings allow for the control, prevention, eradication and elimination of the causes of disease, conditions, injury, disability or death.

viii. Design planning programs, implement, evaluate the program and develop control measures

The systematic use of epidemiological principles and methods for the planning and evaluating of health services is relatively new development. Health services planning are a process of identifying key objectives and choosing among alternative means of achieving them. Thus it helps to control the disease or health related events in the population.

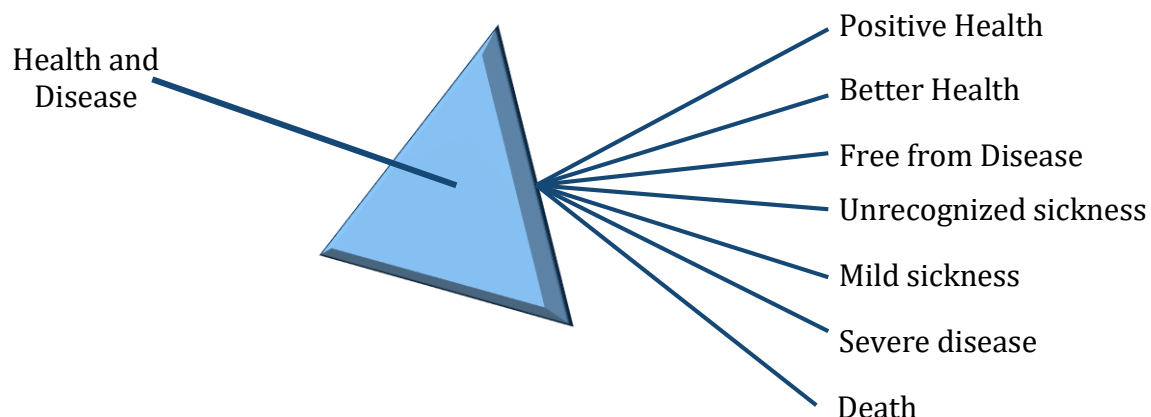
1.9. Contributions of Epidemiology in Medical Field

Epidemiology is the basic science of Public Health and Preventive Medicine. Public health can also be referred as Community Medicine or Population Medicine because primary focus of concern is the community instead of the individuals some of the major contributions of epidemiology in the medical field are as follows;

- Identifying the causes of disease
- Determining the natural history of disease
- Studying the biologic spectrum of disease
- Measuring the extent of disease in a population
- Identifying patterns and trends of disease occurrence
- Improving the diagnosis, treatment, and prognosis of clinical disease
- Determining preventable causes of disease
- Evaluating health interventions at different levels
- Setting disease control priorities in the communities

1.10. Spectrum of health and disease

The spectrum of health and disease is the graphic representation of variation in the manifestation of disease. It is similar to spectrum of light where the colors vary from one end to the other, but difficult to determine where one color ends and the other begins.



The spectral concept of health emphasizes that the health of an individual is not static rather it is a dynamic phenomenon and a process of continuous change with frequent variations. At one end of the spectrum, there are sub-clinical cases which are not ordinarily identified and at the other end there are fatal illnesses. In the middle of the spectrum, there is ranging of illness from mild to severe.

The shifting from optimum health to ill health is often gradual, where one state ends and another begins is a matter of judgment.

Thus, Maximum health considered today may become the minimum for tomorrow. It implies that health is a state not to be attained but ever to be renewed.

1.11. Determinant of Health and Disease

Health is determined by multiple factors. The factors which influence health lie within individual and externally in the environment and surroundings in which the individual lives. The individual factor is determined by genetic factors and the external factors are environmental factors. An individual may contract disease by single factor or the combination of the two or more. These factors interact and these interactions may be health promoting or deleterious. Thus, the health of individuals and whole communities may be considered to be the result of many interactions. Determinants of health deals with health of individual, families, societies and communities.

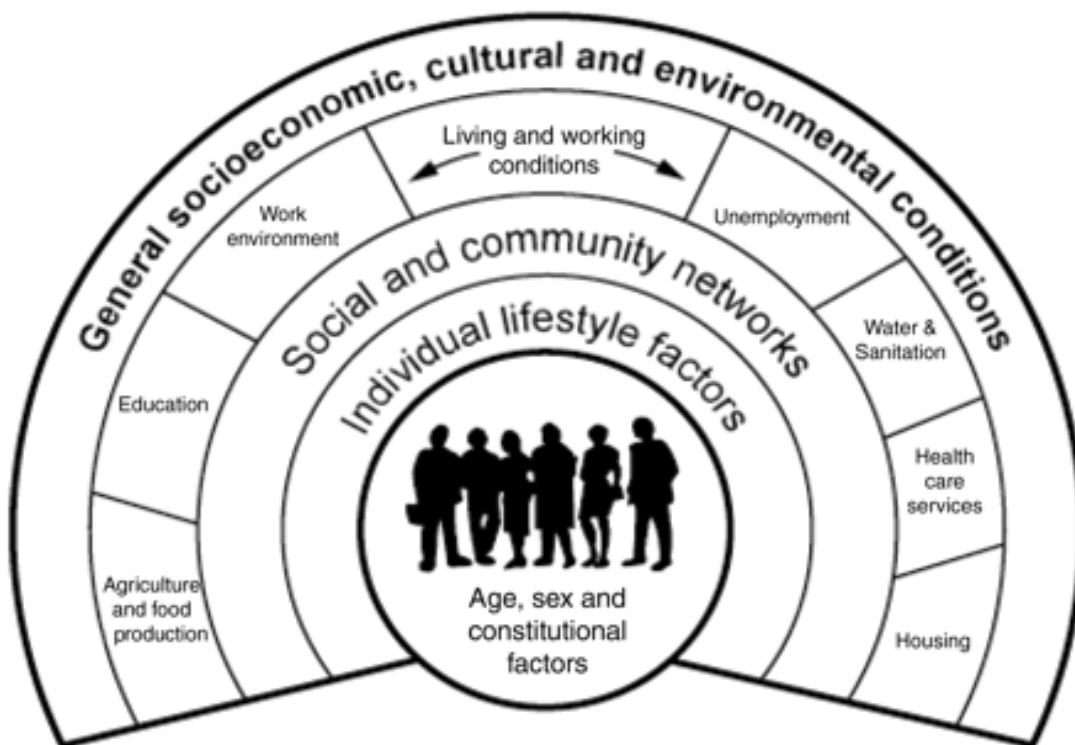


Fig: Determinants of Health and disease

Following are the important determinants of health

- Biological determinants
- Environmental
- Behavioral
- Socio-cultural
- Socio-economic
- Health system
- Ageing of the population
- Science and technology
- Information and communication
- Gender
- Equity and social justice
- Human rights

These determinants are not exhaustive but are suggestive. The complexities in the definition of health, determination of its dimensions further aids in questions that there may be other determinants which is the matter of ever search in epidemiology and public health. An explanation of identified determinants is as follows

1. Biological Determinants:

The physical and mental make up of the individual is determined by the nature of the genes at the moment of the conception. The genetic make up is unique in the sense that it cannot be altered after conception. A number of diseases due to genetic origin are chromosomal anomalies, inborn error of metabolism, mental retardation; some types of diabetes have been recorded. The state of health therefore depends partly on the genetic constitution of an individual.

From genetic view, health may be defined as state of the individual which is based upon the absence from the genetic constitution of such genes as correspond to characters that take the form of serious defect and derangement and to the absence of any aberration in respect of the total amount of chromosome material in the karyotype.

The positive health advocated by WHO implies that a person should be able to express as completely as possible the potentialities of his genetic heritage. The biological determinant of health can be improved through genetic screening and gene therapy.

2. Behavioral and socio-cultural conditions:

The —way people live|| reflects a whole range of social values, attitude and activities, also called life style. It is composed of cultural and behavioral patterns and life long personal habits (e.g., smoking, alcoholism) that have developed through process of socialization. Health is both a consequence of an individual life style and a factor in determining it. Life styles are learnt through social interaction with parents, peer group, friends and siblings and through school and mass media.

Many health problems in developed countries such as Coronary Heart Diseases, obesity, lung cancer, drug addiction are associated with life style changes. Similarly in developing countries

risk of illness and death are connected with lack of sanitation, poor nutrition, personal hygiene, customs and cultural patterns. Achievement of optimum health requires adoption and promotion of healthy lifestyles.

3. Environment:

The concept of disease-environment association was related by Hippocrates which was later on revived by Pettenkofer. Environmental determinants are classified into internal and external factors. The internal environment includes each and every component part, every tissue, organ and organ system, and their harmonious functioning within the system.

The external environment consists of those things to which man is exposed after conception. It is defined as –all that is external to the individual human host. The external factors can be divided into physical, biological and psychosocial components which are health promoting or detrimental to health of an individual.

The environmental factors range from housing, water supply, psychosocial stress, and family structure through social and economic support system, to the organization of health and social welfare services in the community. If the environment is favorable to the individual, he can make full use of his physical and mental capabilities. Hence, protection and promotion of family and environmental health is important for better health.

4. Socio-economic conditions:

Socio-economic conditions have long been known to influence human health. Health status is determined primarily by their level of socioeconomic status e.g. per capita GNP, education, nutrition, employment, housing, political system of a country, etc.

The components of socioeconomic system are as follows:

a. Economic status:

The per capita GNP is most widely accepted measure of general economic performance. There can be no doubt that in many developing countries, it is the economic progress that has been the major factor in reducing morbidity, increasing life expectancy and improving the quality of life. The economic status determines the purchasing of power, standard of living, quality of life, family size and the pattern of disease and behavior in the community. It is also an important factor in seeking health care. Ironically, affluence may also be a contributory cause of illness as exemplified by the high rates of coronary heart disease, diabetes and obesity in upper socioeconomic groups.

b. Education/ literacy status

A second major factor influencing health status is education (especially female education). The world map of illiteracy closely coincides with the maps of poverty, malnutrition, ill health, high infant and child mortality rates. Studies indicate that education, to some extent, compensates the effects of poverty on health, irrespective of the availability of health facilities.

c. Occupation:

The very state of being employed in productive work promotes health, because the unemployed usually show a higher incidence of ill-health and deaths. For many, loss of work may mean loss in income and their status. It can cause psychological and social damage.

d. Political system:

Health is also related to the country's political system. Often the main obstacles to the implementation of health technologies are not technical, but rather political. Decisions concerning resource allocation, manpower policy, choice of technology and degree to which health services are made available and accessible to different segments of the society are examples of the manner in which the political system can shape community health services. The percentage of GDP spent on health is a quantitative indicator of political commitment.

To achieve the goal of health for all, WHO has set the target of at least 5 percent expenditure of each country's GDP on health care. What is needed is political commitment and leadership which is oriented towards social development, and not merely economical development. Social, economic and political actions are required to eliminate health hazards in people's working and living environments.

5. Health services

The term health and family welfare services cover a wide spectrum of personal and community services for treatment of diseases, prevention of disease and promotion of health. The purpose of health services is to improve the health status of population. For example, immunization of children can influence the incidence/prevalence of particular disease. Provision of safe water can prevent mortality and morbidity from water-borne diseases.

The care of pregnant women and children will contribute to the reduction of maternal and child morbidity and mortality. To be effective, the health services must reach the social periphery, equitably distributed, acceptable, adequate, appropriate, comprehensive, feasible and affordable.

All these are ingredients what is now termed —primary health care||, which is seen as the way to better health. Health services can also be seen as essential for social and economic development. The most we can expect from an effective health service is good care. The epidemiological perceptive emphasizes that health services, no matter how technically cost-effective, are ultimately pertinent only if they improve health.

6. Aging of the population:

By the year 2020, the world will have more than one billion people aged sixty or over and more than two-thirds of them living in developing countries. A major concern of rapid population aging is the increased prevalence of chronic diseases and disabilities both being condition that tend to accompany the aging process and deserve special attention.

7. Gender

The 1990's have witnessed an increased concentration on women issues. In 1993, the Global commission on women's health was established. The commission drew up an agenda for action on women's health covering nutrition, reproduction, health consequences of violence, aging, lifestyle related conditions and the occupational environment. It has brought about an increased awareness among policy-makers of women's health issues and encourages their inclusion in all development plans as a priority.

8. Other factors

We are witnessing the transition from post industrial age to an information age and experiencing the early days of two interconnected revolutions, in information and in communication. The development of these technologies offers tremendous opportunities in providing an easy and instant access to medical. It contributes to dissemination of information worldwide, serving the needs of many physicians, health professionals, biomedical scientists and researchers and other people.

Other contributions to the health of population derive from systems outside the formal healthcare system. i.e., health related systems (e.g., food and agricultural, education, industry, social welfare, rural development) as well as adoption of policies in economical and social fields that would assist in raising the standards of living.

This would include employment opportunities, increased wages, prepaid medical programmes and family support systems. In short, medicine is not the sole contributor to the health and well being of population. The potential of inter-sectoral contributions to the health of communities is increasingly recognized.

1.12. Basic Epidemiological Measurements

Inherent to the definition of epidemiology, it is the measurement of various phenomena relating to human life and health problems. These phenomena can be expressed in terms of rate, ratio and proportion.

1. Rate

The basic of epidemiology is to assess the amount of growth, disease, disability, injury and death in the populations. It is also equally important to study and analyze any or all factors affecting these measures of the health status. It is also important to present the data and statistics in a manner that makes sense and is comparable with one another. Simply, expressing the variables in percentage is not sufficient to compare the health status of various communities. This does not provide the degree of severity of the health events or disease in real sense for the epidemiologist. Thus, a rate is needed to compare with to another.

A rate is the number of frequency of a disease per unit size of a population of group. The unit size is presented as per 100. 1000. 10000. 100000.

A rate measures the amount of disease, event, condition, injury, disabilities and/or death in a population, group, community or geographical area by relating cases of the disease to the population base.

A basic formula for a rate includes following elements.

- Numerator: It denotes the number of cases with the disease
- Denominator: It denotes the population of the given area or risk population
- Time period: It denotes the specific time period that the events occur.
- Multiplier: As multiplier or constant (10^n)

Formula:

$$\text{Rate} = \frac{\text{Number of cases} \times 10^n}{\text{Population of the given area in the specified time period}}$$

The numerator is confined to a specific set of characteristics such as age, sex, race, occupation, religion and education etc. The denominator is limited to the population of the study group or the total population such as city, school, community, region, state etc.

High rates as well as low rates provide useful information and insights into cause, spreads, transmission, and overall effects of a disease on a population.

Types of Rate

There are three types of rates. They are given below.

- a. **Crude rates:** These rates are based on the number of experiences or events that happen in a total population in a certain time period. Two crude rates are fundamental to epidemiological methods. Eg. CDR, CBR.
- b. **Adjusted rates/Standardized rate:** It is important to evaluate specific rates and age distributions in order to ascertain if an adjustment is needed, when assessing crude rates. These use mathematical calculations and transformations to allow comparisons among and between populations having characteristics or traits that differ or which may affect risk of injury, disease, disability or death.
- c. **Specific rates:** A specific rate provides detailed information presented as rates for specific age, religion, race, gender etc. The denominator of a specific rate uses a specific population or sub group of population for a certain geographic area and certain period.

2. Ratio

The ratio is the relation in number, degree, or quantity existing between two things; a fixed relationship in number or degree between the two similar things is existed. Ratio is the result of one quantity divided by another of the same kind and expressed as a fraction. In a ratio, the numerator is not included in the denominator. It is a relative number that expresses the magnitude of one in relation to another variable.

$$\text{Ratio} = \frac{\text{One quantity (X)}}{\text{Another quantity (Y)}}$$

Or X:Y

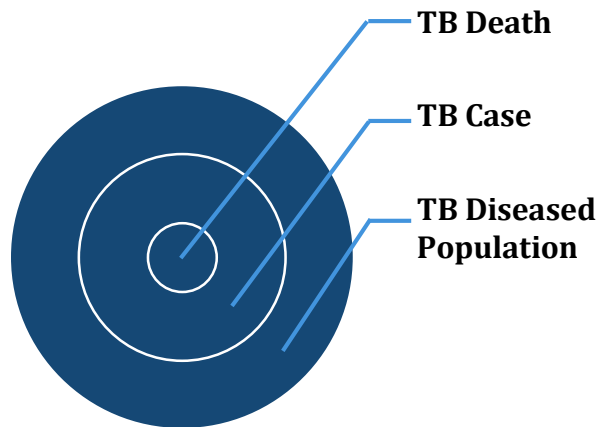
Example: Male- female ratio, Doctor - population ratio, patient - Bed ratio

3. Proportion

A proportion is the relation between the amount, number, size, or degree of one thing and the amount, number, size, or degree of another. In epidemiology, of the numbers of those people currently with a disease or condition are expressed relative to the total number of

all who have had the disease or condition, then this is known as proportion. If it expressed in relation to total number of population then it called rate.

Proportion is a ratio in which the numerator is included as a part of the denominator. The value of proportion falls within the range of 0.0 to 1.0. The important difference between a ratio and proportion is that the numerator of a proportion is included in the population defined by the denominator. Some example are Proportional mortality, case fatality etc. Example (Illustration of proportion)



The scope of measurements in epidemiology is very broad and unbounded. It includes following measurements

- A. Measurement of mortality
- B. Measurement of morbidity
- C. Measurement of disability
- D. Measurement of fertility
- E. Measurement of attributes of disease
- F. Measurements of health care needs and health care utilization of people
- G. Measurement of demographic variables

The basic requirements of measurements are validity, reliability, accuracy, sensitivity, and specificity. A short description of each measurement is as follows (only mortality and morbidity indicators are described here)

(A) Measurements of mortality

Basic measurements of mortalities are:

1. Crude death rate

It is defined as the number of deaths of people irrespective of age, sex and cause, per thousand populations per year in a given community. It indicates the rate at which people are dying.

$$\text{CDR} = \frac{\text{Total number of deaths in given time period (usually one year)}}{\text{Mid year population of given area}} \times 1000$$

2. Infant mortality rate

Infant mortality rate is defined as the ratio of deaths of children below one year of age to the total number of live births of the same year.

$$\text{CMR} = \frac{\text{Total number of deaths of children below 1 years of age in a given area in defined period of time}}{\text{Total number of live birth of same year in given area}} \times 1000$$

Note:

- IMR is commonly used indicator of health status of health
- It is most sensitive indicator of health
- High rate of IMR signifies low access and coverage of health services.

3. Child mortality

Child mortality is defined as the number of deaths at age 1-4 years in a given year, per thousand children in the same age group at the midpoint of the year. Mathematically child mortality is calculated as:

$$\text{CMR} = \frac{\text{Total number of deaths of children of age group 1-4 years in a given area in defined period of time}}{\text{Total number of children of the same age group in given area}} \times 1000$$

High rate of child mortality signifies there is insufficient nutrition, low coverage of immunization services and adverse environmental exposure factors.

4. U-5 five mortality

Under five mortality is defined as the total number of deaths of children below five years in the given year per thousand live births of same year in specified area.

$$\text{U5MR} = \frac{\text{Total number of deaths of children below five years in given area in defined period of time}}{\text{Total number of live birth of the same year}} \times 1000$$

High rate of under five mortality reflects high birth rate, high mortality and shorter life expectancy.

5. Maternal mortality rate

Maternal mortality is defined as the death of woman while pregnant or within 42 days of termination of pregnancy, irrespective of duration and site of pregnancy, from any cause related to or aggravated by pregnancy or its management but not from accidental or incidental cause - WHO

$$\text{MMR} = \frac{\text{Total number of female deaths due to complication of pregnancy, child birth or within 42 days of termination of pregnancy from puerperal causes in an area during a given year}}{\text{Total number of live birth of the same year}} \times 1000$$

6. Specific Death rate

It measures the rate of dying due to particular cause or attribute (Age, sex, conditions etc). Some examples include deaths of tuberculosis, cholera deaths it is calculated as:

$$\text{SDR of TB} = \frac{\text{Total number of Deaths of people due to tuberculosis in an area}}{\text{Total population of the area in given time}} \times 1000$$

7. Life expectancy

Life expectancy at birth is —the average numbers of years of an individual of a given age is expected to live if current mortality rates continue||. Life expectancy at birth is highly influenced by infant mortality. An increase in life expectancy is regarded as the improvement in the health status of population.

Uses of Mortality Data

1. Mortality data are to describe the trend of deaths
2. To identify the causes of deaths
3. Provides the basis of prioritization and programming
4. Monitoring and evaluation of intervention/health status of people.
5. Comparison of health status of different communities/countries
6. Gives clue on epidemiological researches.

(B) Measurement of Morbidity

The principle of measurement of frequency of disease is based on the clear definition of what is to be measured and establishment of criteria or standard by which it can be measured. The definition of disease frequency measurement should be precise; valid to enable to identify those who have diseases from those who do not, acceptable and applicable to its use in large populations.

Clear definition may help to minimize error in classification of data and standardize the methods of observation and recording. Several measure of disease frequency is based on the fundamental concept of prevalence and incidence.

The basically measures of disease includes three different methods. These include

1. Measure of disease frequency
2. Measure of disease association
3. Measure of impact of disease

1. Measure of disease frequency

Principle methods of measurements of frequency of disease include prevalence and incidence. These are counted as the number of diseased person, disabilities, episodes of illness, attack rates and defining susceptible population etc. Investigations of these conditions fall in the domain of descriptive epidemiology and causal investigation.

2. Measure of disease association

This aspect deals with the measurement of relationship between causes and potential outcomes. Some methods include estimation of odds ratio, risk difference, relative risk, and defining specific casual criteria for the establishment of association etc. Statistics is the tool in making decision. If there is statistical relationship, then there may be significant association between cause and effects.

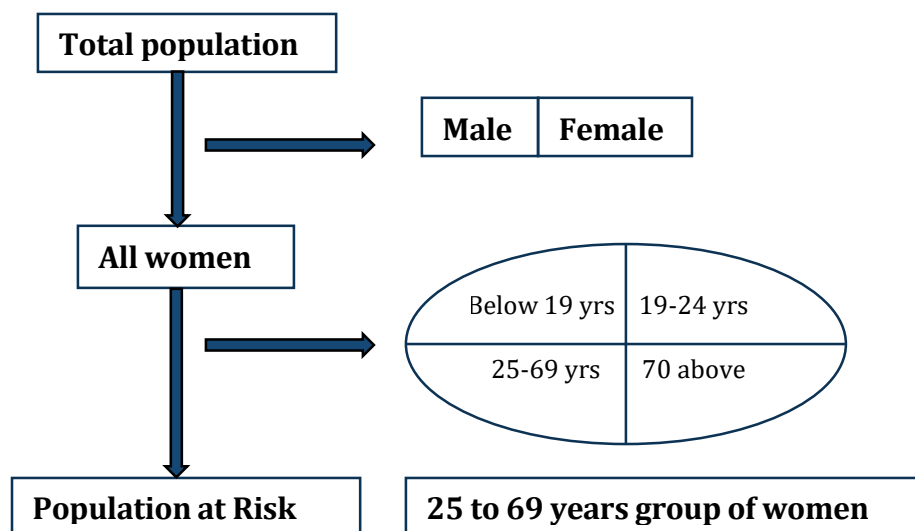
3. Measure of impact of disease

These reflect the expected contribution of factors leading to the problems and its outcomes. These measures are useful in estimation of efficacy and effectiveness of

intervention. Assessment of potential impacts acts as useful guideline formulation and development of relevant strategy for health promotion, reduction of impact of disease and way to forward in future.

Concept of population at risk

That part of population which is susceptible to a disease is called the population at risk. It can be defined on the basis of demographic or environmental factors. For example, occupational injuries occur only among factory/industry workers, females over 25 years are at risk of developing carcinoma of cervix.



Morbidity: Morbidity is an extent of illness, injury, disorder, sickness, disease, illness and disability in a defined population. It is also any deviation from a state of health and wellbeing or presence of a morbid condition. Morbidity is expressed in general or specific rates of incidence and prevalence. It also refers to sickness rates; the number of sick persons compared to the certain population which is often well or at risk group. In epidemiology, the main measures of morbidity are incidence and prevalence rates and the various sub measure of each. Morbidity could be measured in terms of 3 units.

- Persons who are ill
- The illness (periods or spells of illness) that these persons experienced
- The duration (hour, days, weeks etc) of these illness

Different types of frequency measures of diseases

1. Incidence rate
2. Attack rate (primary and secondary attack rate)
3. Case fatality rate
4. Prevalence rate

1. Incidence and Incidence Rates

Incidence is used as a measure of the rate at which new cases of disease, disorder, injury or health related event occur in a population. Incidence is the number of new cases of a disease

which came into existence within a certain period of time per specified unit of population. In other way, it is defined as the number of new cases of a disease occurring in a specific population in a specified time period.

$$\text{Incidence rate} = \frac{\text{Number of new cases within a population in a given time period}}{\text{Number of persons exposed to risk of developing the disease in the same time period}} * 10^n$$

(Where n is the multiplying factor = 1, 2, 3 ...)

Incidence of disease can be measured by two methods.

- Person - time incidence (incidence rate or incidence density)
- Cumulative incidence (incidence risk)

Person - time incidence (incidence rate or incidence density)

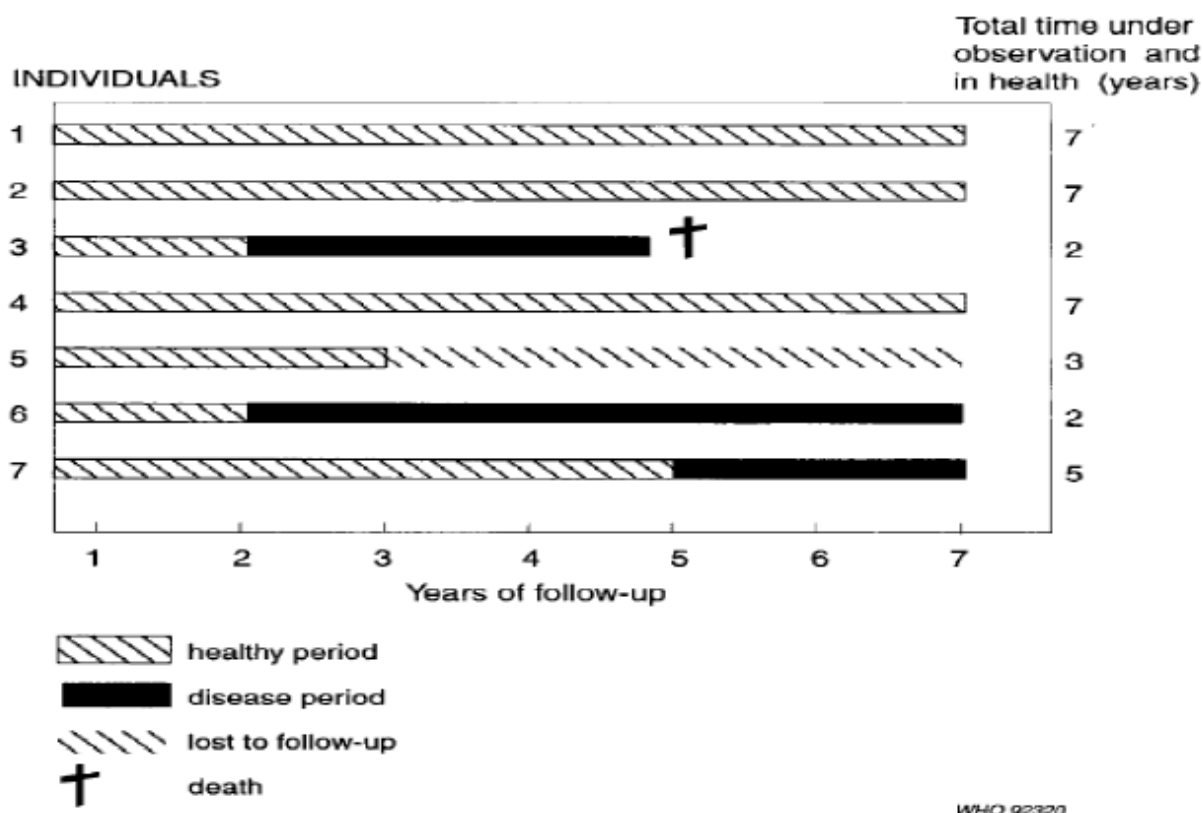
It is also called incidence rate or incidence density. It refers to the number of new cases of disease or events that occur in a defined time period and the denominator is the population at risk of experiencing the events during this period. The most accurate way of calculating the incidence rate is the to calculate person-time incidence rate. Each person in the study population contributes one person- year to the denominator for each year of observation before disease develops or the person is lost to follow up. Mathematically it is calculated as

$$\frac{\text{Number of people who get a disease in a specified period of time}}{\text{Sum of length of time during which each person in the population is at risk.}} * 10^n$$

Essential features of Incidence rate

- The numerator strictly refers only to the first events of disease.
- The unit of incidence rate must always include a dimension of time (day, week or year)
- each individual during the observation should be disease free
- denominator will be the sum of all disease free time periods in the defined time period of study
- It can also refer new spells or episodes of disease arising in a given time period

Example



Here

- Total numbers of disease free people at the beginning of study = 7.
- Total numbers of events (diseased people) = 3
- Sum of length of time during which each individual is at risk of developing disease = (7+7+2+7+3+2+5= 33)

Incidence rate = total number of cases / total population at risk (sum of all risk) $3/33 \times 100 = 9.1$ cases per 100 person year

Cumulative incidence

It is also known as incidence risk. It is a simpler measure of the occurrence of disease or health status. It measures the denominator at the beginning of study. It is given by:

$$CI = \frac{\text{Number of people who develop disease in particular time}}{\text{Total number of disease free people at the beginning of study}} \times 10n$$

Let us consider the values of above example

$$CI = 3/7 \times 100 = 43 \text{ cases per 100 persons during the seven years.}$$

Special incidence rates

a. Attack rate:

It is an incidence rate used when the population is exposed to risk for a limited period of time such as during an epidemic. It is usually expressed in percent. It relates the number of cases in the population at risk and reflects the extent of the epidemic

For a defined population (the population at risk) during a limited time period,

$$\text{Attack rate} = \frac{\text{New cases occurring during a given time period}}{\text{Population at risk during the same time period}} \times 100$$

Example

Among 75 persons who attended a church picnic, 46 subsequently developed gastroenteritis.

- Cases of gastroenteritis occurring within the incubation period for gastroenteritis among persons who attended the picnic = 46
- Number of persons at the picnic = 75

Then, the attack rate for gastroenteritis is $46/75 \times 100 = 61\%$

This is a measure of the probability or risk of becoming a case.

In the example above,

Probability of developing gastroenteritis was 61%,

Or the risk of developing gastroenteritis was 61%.

b. Secondary Attack Rate

A secondary attack rate is a measure of the frequency of new cases of a disease among the contacts of known cases. The formula is as follows:

$$\text{Secondary attack rate} = \frac{\text{No. of cases among contacts of primary cases during the period}}{\text{Total number of contacts}} \times 100$$

To calculate the total number of household contacts, we usually subtract the number of primary cases from the total number of people residing in those households.

Example

Seven cases of Hepatitis A occurred among 70 children attending a child care center. Each infected child came from a different family. The total numbers of persons in the 7 affected families were 32. One incubation period later, 5 family members of the 7 infected children also developed hepatitis A. Calculate the attack rate in the child care center and the secondary attack rate among family contacts of those cases.

1. Attack rate in child care center:

Cases of hepatitis A among children in child care center = 7
Number of children enrolled in the child care center = 70
Attack rate = $7/70 \times 100 = 10\%$

2. Secondary attack rate:

Cases of hepatitis A among family contacts of children with hepatitis A = 5
Number of persons at risk in the families (total number of family members - children already infected) = $32 - 7 = 25$
Secondary attack rate = $5/25 \times 100 = 20\%$

c. Case fatality Rate

Case fatality is a measure of severity of a disease and is defined as the proportion of cases of a specified disease or condition which are fatal within a specified time.

$$\text{CFR} = \frac{\text{Number of deaths from a disease in a specified period}}{\text{Number of diagnosed cases of the disease in the same period}} \times 100$$

Uses of incidence rate

- It is useful for taking action to control disease
- It is useful for research the determinants (etiology) and pathogenesis of disease.
- It is useful to know the distribution frequency of disease.
- It helps to develop the efficacy of preventive and therapeutics measures

(C) Prevalence and prevalence rates

Prevalence is the number of cases of disease, infected persons, or condition, present at a particular time, in relation to size of the population from which it drawn. It includes both new and old cases (current cases) occurring in the population.

Prevalence rates

The prevalence rate refers to all current cases (old and new) existing in given point in time, for over a given period of time in a given population. In broader sense, it is defined as the total number of all individuals who have an attribute or disease at a particular time/during particular time period divided by population at risk of having the attribute or disease at same point of time or midway of time period.

$$\text{Prevalence rate} = \frac{\text{Total number of cases of the disease at a given time}}{\text{Total population at risk at a given time}} * \text{Multiplier}$$

$$\text{Multiplier} = 10 \text{ n or } 1000$$

The prevalence rate assesses the total number of people in a group or population who have a disease at a specific time. The prevalence rate is equal to the incidence rate multiplied by the average duration of the disease. Prevalence is controlled by two elements.

- 1) The number of individuals who have been diseased in the past
- 2) The length or duration of illness.

Prevalence is varying in direct relation to both duration and incidence. Successful treatment and intervention will prolong life and will have an effect on reducing prevalence. The decrease in incidence and shortening the duration of a disease decreases prevalence. When duration decreases significantly prevalence can decrease despite an increase in incidence. Prevalence rates and information obtained can be used in planning the programme. The formula for prevalence rate is given below.

a. **Period prevalence rate**

Period prevalence includes the total individuals who have had the disease of concern at any time during specified time period. It starts at a point in time and stops at a point in time. All persons with disease that have carried over from the previous time or have become ill at the end of the time period are included. It includes point prevalence at the beginning of the time period plus all new cases that occur within the time period. Period prevalence requires the establishment of a specific time period for the study of a specific disease. Period prevalence is a complex measure. It is established from the prevalence at a point in time plus incidence (new cases) and recurrences during a succeeding time period such as a one year. Persons not at risk of acquiring a new case of disease are not included in the numerator. Below is the formula of period prevalence.

$$\text{Period prevalence} = \frac{\text{Number of existing cases of the disease with in a time period}}{\text{Average study population}} * 1000$$

b. Point prevalence rate

It is the number of cases of individuals with a disease, condition, or illness at a single specific in time; the number of existing cases at a point in time. Point prevalence measures the presence of a disease or condition on a single short time point. Theoretically point of time indicates stopping the clock for a minute, hour and or a day and counting the existing cases of disease. Point prevalence rate is given below.

$$\text{Point prevalence} = \frac{\text{Number of existing cases of the disease at a point in time}}{\text{Total number of study population}} * 10n$$

Comparative factors affecting prevalence rates

Rates are increased by	Rates are decreased by
– Immigration of ill cases – Emigration of healthy persons – Immigration of susceptible cases or those with potential of becoming cases – Prolongation of life of cases without cure (increase the duration of disease) – Increase in occurrence of new cases (increase in incidence)	– Immigration of ill cases – Emigration of healthy persons – Improved cure rates of cases – Increased death rates from the disease. – Decrease in occurrence of new cases, shorter duration of disease and death

In addition to these, other useful mathematical indicators available to assess disease problem include notification rate, attendance rate at OPD/IPD, hospital admission rate and discharge rate.

Relationship between incidence and prevalence

Prevalence rate of disease depends on various factors. It relies on incidence and the duration of particular disease. Prevalence of disease is the product of duration and incidence rate. Let P be the prevalence, I be the incidence and duration will be D, then

$$P = I \times D$$

Uses of Morbidity data

1. Describes the magnitude and nature of health problems
2. Useful to describe health status of people
3. Helps to define population at risk.
4. These are useful for administrative and planning purposes.
5. These can be used as the parameters for comparison of health status at local, national and international level. Thus, can be used advocacy tool.
6. Provide more comprehensive and relevant information regarding behavioral pattern and characteristics of patient.
7. Disclose the natural history of disease and syndromes to define the disease

CHAPTER 2

Epidemiological Methodology/Studies (Study Design)

Primary concern of epidemiologist is to study the disease occurrence among people who during the course of their lives are exposed to numerous factors, circumstances and conditions, some of which may have role in disease etiology. Epidemiologist employ carefully designed research strategies to explore disease etiology. Thus, Epidemiological studies are concerned with the natural history and etiology of the diseases.

2.1 Classification of epidemiological study designs

Epidemiological studies can be classified into various types on several bases. Some types

1. Classification of epidemiological studies on the basis of philosophy adopted

Epidemiological studies are classified into conceptual and empirical type. Conceptual studies are related to some abstract ideas or theory and used by philosophers and thinkers to develop new concepts or to reinterpret existing ones. An empirical study relies on the experiences or observation. It is data based, coming up with conclusions which are capable of being verified by observation or experiments. It can also be called as experimental study.

2. Classification of epidemiological studies on the basis of type and aspects of the study.

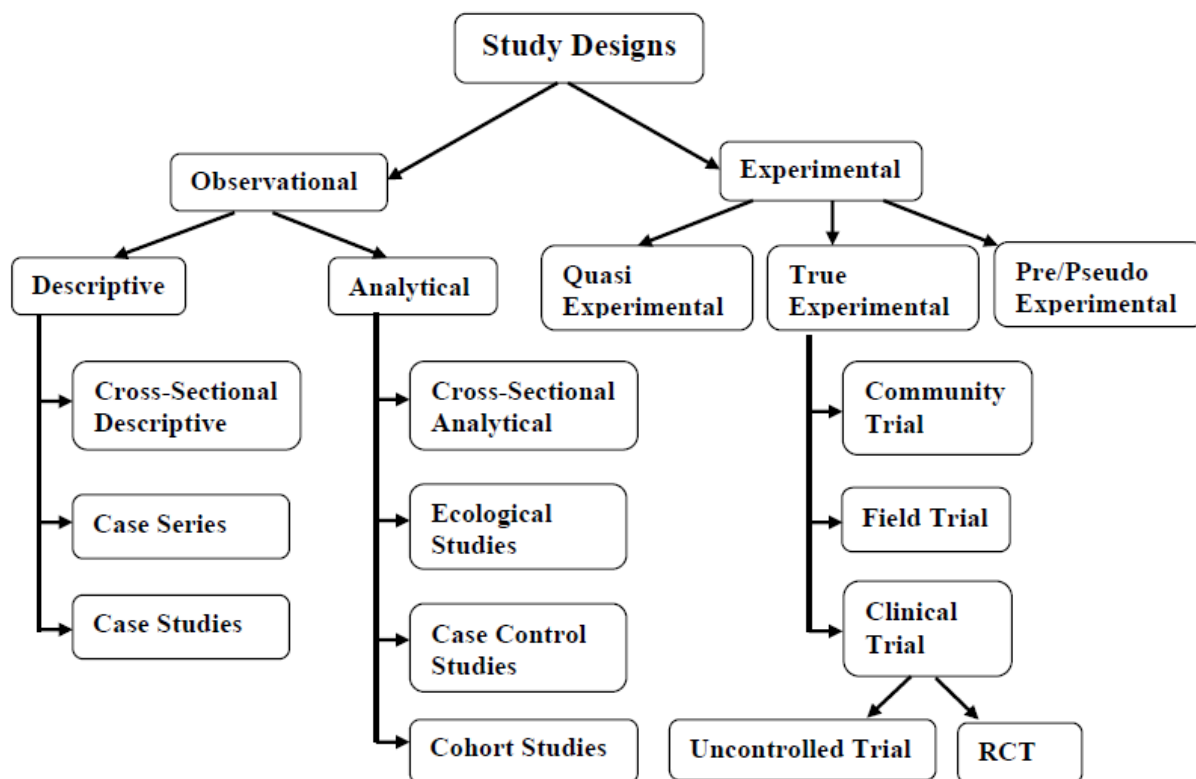
Epidemiological studies can be classified as either observational or experimental. Observational studies allow nature to take its course; the investigator measures but not intervenes. Observational studies have further subdivisions: Descriptive and Analytical. A descriptive study is limited to description of occurrence of a disease in a population and is often the first step in epidemiological investigation. An analytical study goes further investigation by analyzing the relationship between health status and other variables. Unlike observational studies, in experimental studies the researcher describes and manipulates the objects or situations either by preventive or therapeutic measures and determines the outcome of his manipulation. Experimental studies are also subdivided into clinical trial, community trial and field trial.

3. Classification of epidemiological studies on the basis time of study Epidemiological studies may be either **cross- sectional or longitudinal**. Cross- sectional studies are done at one point in time while longitudinal studies are performed over a prolonged period of time by follow up examinations.

4. Classification on of epidemiological studies on the basis of intervention on natural course

- Non intervention study: In this study, the researcher just describes and analyses events and draws the inferences. No intervention takes place.
- Intervention study: the researcher describes and manipulates the objectives or situations and look over action of these on the outcomes.

The extended form of classification of epidemiological studies that include almost all aspects is as follows:



1. Observational study

- a. Descriptive study: Case report and Case Series
- b. Analytical study
 - i. Cross sectional study/Prevalence study
 - ii. Ecological Study/Correlation study
 - iii. Longitudinal Study
 - (1) Case control study/Backward looking study/Retropective study
 - (2) Cohort study/Forward looking study/Prospective study/Incidence study

2. Experimental study

- a. Randomized Control Trial
- b. Field Trial
- c. Community Trial

A) Observational study/non intervention study

Observational study is one in which the nature takes its own course. The investigator only measures the effects but does not intervene anywhere in the whole process.

a. Descriptive study

Descriptive study is limited only to a description of the occurrence of a disease or an event in a population.

b. Analytical study

Analytical study is one in which analysis is done between two variables.

1. Cross sectional

It is based on a single observation on a cross section of population at one point of time. It is useful in case of chronic diseases. It is also called 'Prevalence study'.

2. Ecological study

Ecological study is one in which study done in one ecological zone is compared with other to establish relationship between cause and effect. It is based on the study of events in relation to particular attribute in different or same group.

3. Longitudinal

It is based on the repeated observation in the same population over a prolonged period of time by means of follow up examinations.

i) Case control Study

It is based on the comparison between the cases and control.

ii) Cohort Study

Effects are observed among the cohort for a period of time. Cohort is a group of people who share a common characteristic, e.g., Birth cohort, Marriage cohort, Exposure cohort, etc.

B) Experimental / Intervention study

Experimental study is one in which the change is observed after an intervention.

1. Randomized Controlled Trial

It is done among two different randomly selected groups of study elements. It is usually done in hospitals.

2. Field trial

It is done in randomly selected groups of people in the community.

3. Community Trial It is performed in a community with similar culture and other variables. It is particularly useful where the problem is of social origin having health consequences. **1.**

2.2 Descriptive study

Descriptive study is the basis of other epidemiological studies. Some times this study is required to find out the hypothesis of analytic and experimental studies. Prior to the start of the study/investigation of an occurrence, some logical questions arise. These questions are;

1. When it occurs?
2. Where is occurs?
3. Who are affected?

Study design/steps of Descriptive study

1. Defining the population to be studied
2. Defining the occurrence/disease under study

3. Describing the occurrence/disease by ; Time, Place and person
4. Measurement of the occurrence/disease
5. Comparing with known indices
6. Formulation of hypothesis

1. Defining the population

The first step of the study is to define the population under study (population at risk), in terms of;

- Total number
- Age
- Sex
- Occupation
- Culture, etc.

2. Defining the disease under study The next step is to define or specify the disease being studied / investigated. The researcher must identify the disease/occurrence in question and he must spell out clearly the criteria by which the disease can be measured (operational definition).

3. Describing the disease The primary objective of descriptive study is:

- To describe the occurrence and distribution of the disease by (a) time, (b) place and (c) person
- To identify those characteristics associated with the presence and absence of the disease among the population under study.

This involves the systematic collection and analysis of data.

(a) Time distribution

The occurrence of disease changes over time. Some of these changes occur regularly, while others are unpredictable. Such pattern of disease may be described by the time of its occurrence i.e., by weeks, months, years, the day of the week, hour of the onset, etc. There are three types of time trend;

1. Short term fluctuations: The best example of short term fluctuation is an Epidemic: Epidemic may be defined as the number of occurrence of any health related event in a population is clearly an excess of its normal expectancy. Epidemic can be classified as:

a. Common source epidemic

- Single exposure epidemic (Food poisoning)
- Multiple exposure epidemic (VD from prostitute)

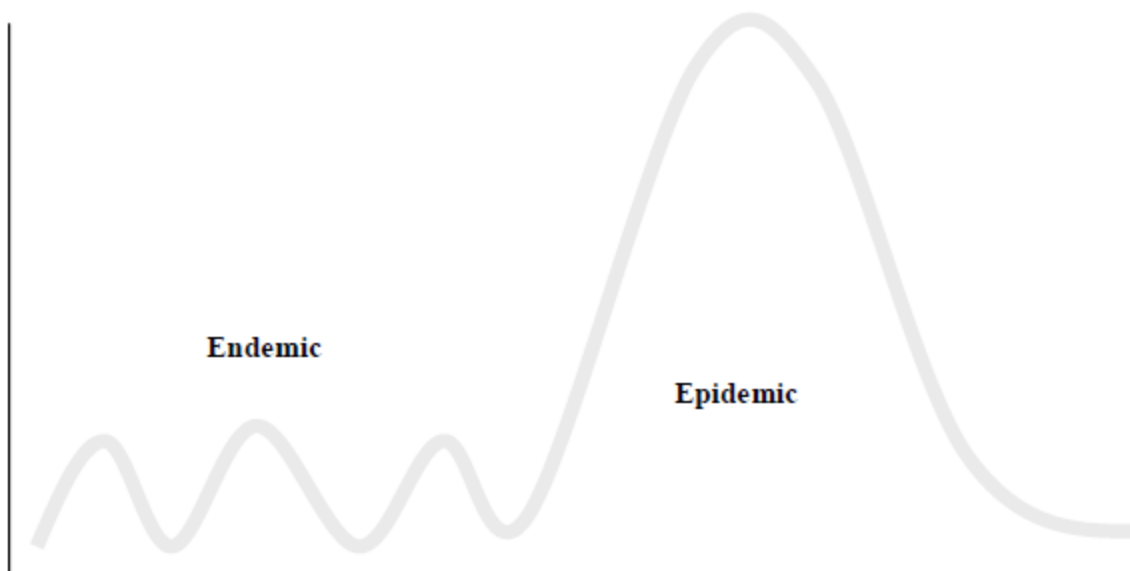


Figure: Point source (Single source) epidemic

b. Propagated source epidemic

- Person to person/Direct contact (Polio)
- Vector borne (Malaria)

c. Modern or slow epidemic:

2. Periodic fluctuations or seasonal Variation

These may be seasonal variation or cyclic variation. For example; Early spring – Measles, Chickenpox, etc. Winter – URTI, Frostbite, etc. Summer – GIT infections, Sun strokes, snake bite, etc. Figure: seasonal variation Rubella , Influenza, and rotavirus

3. Long term or secular trend

It is the change of occurrence of diseases (increase or decrease) over a long period of time, e.g., IHD, Lung cancer; Diabetes mellitus are on increase in our country.

(b) Place distribution (Geographic variation)

It provides an important source of clues about the causes of diseases or events. Geographic variation may be classified as;

- a. National variation
- b. Rural-Urban variation
- c. Local variation
- d. **International variation** The pattern of diseases is not the same everywhere in the world. e.g., Cancer of stomach is very common in Japan but unusual in US. Cancer of cervix is common in Indo-Pak-Bangla sub-continent but

uncommon in industrialized countries. Breast cancer is more prevalent in western countries but less prevalent in Japan.

(c) Person distribution

Defining the persons who develop the diseases is the most important factor in descriptive study. These are Age, Sex, Occupation, Marital status, Habits, Social class, Religion, culture, custom, etc.

- **Age:** Age is an important factor for variation of disease prevalence. For example; Measles is more common in children. Cancer in middle ages and Atherosclerosis in old ages.
- **Sex:** Diabetes, Hypertension, is more common and lung cancer, coronary heart diseases are less common in females.
- **Marital Status:** Mortality rate is lower in married than un-married persons of the same age and sex.
- **Occupation:** Coal miners suffer from Silicosis, sedentary workers from IHD and farmers from tetanus and hook worm.
- **Socioeconomic status:** Socioeconomic status is difficult to quantify. It is made up of many variables such as occupation, family income, educational achievement or census track, living conditions, and social standing.

Epidemiologists commonly use occupation, family income, and educational achievement, while recognizing that these variables do not measure socioeconomic status precisely. The frequency of many adverse health conditions increases with decreasing socioeconomic status. Similarly a few adverse health conditions occur more frequently among persons of higher socioeconomic status.

Example Tuberculosis is more common among people with lower socioeconomic strata. Infant mortality and time lost from work due to disability are both associated with lower income. People of lower social class suffer from deficiency diseases, infectious diseases, etc. People of upper social class suffer more from many chronic diseases. Gout was known as the —disease of kings|| because of its association with consumption of rich foods. Other examples are IHD, Hypertension, Obesity, etc.

- **Behavior:** Change of behavior affects the individual for causation of some specific diseases, like Smoking causes lung cancer, heart diseases, Drug abuse causes HIV/AIDS, HBV infection and over eating causes obesity, etc.
- **Ethnic and racial groups.** Sometimes epidemiologists are interested in analyzing person data by biologic, cultural or social groupings such as race, nationality, religion, or social groups such as tribes and other geographically or socially isolated groups. Differences in racial, ethnic, or other group variables may reflect differences in susceptibility or exposure, or differences in other factors that influence the risk of disease

4. Measurement of disease

It is mandatory to have a clear load of disease in the population. This information should be available in terms of mortality, morbidity, disability rates.

5. Comparing with known indices

The essence of epidemiology is to make comparison and ask questions. By making comparison with different groups of population sub groups, it is possible to come at conclusion that whether the identified facts are supportive to the established facts or not.

6. Formulation of Hypothesis

By studying the pattern of distribution of disease or health events, it is possible to formulate hypothesis relating to disease etiology. A hypothesis is a supposition, arrived from observation. It is accepted or rejected using the techniques of analytical epidemiology.

Use of results of Descriptive Study The results of descriptive study provide;

1. Information on the magnitude of the problem/disease in the community.
2. Information on the type of disease/problem in the community in terms of morbidity and mortality rates and ratios.
3. Clues to disease etiology and helps in the formulation of hypothesis.
4. Background information/data for planning, organizing, and evaluating preventive and curative services.
5. Basic data (base line information) for future research.

Case studies or case reports and case series

are the best examples of descriptive studies. Studying an individual for a short span who presents his/her experiences, whereby some new or atypical information can be realized. The study can be conducted in groups of population where several individuals are observed for any new information come out from the case report. Information may lead to formulate new hypothesis. This is a common form of study in clinical setting. A case study includes the intensive information regarding a case or cases whereas case series represents the compilation of similar case reports. Conclusion can be drawn from there reports to lead to formulate hypothesis.

Descriptive vs. Analytical study

Descriptive	Analytical
Describes the disease occurrence with general information	Explains the disease occurrence
Distribution of disease	Determinants of disease
More diffuse & superficial.	Narrow down to answer specific questions
No attempt to analyze the link between exposure and outcome	Exposure and outcome relationship is analyzed
Usually no hypothesis testing	Hypothesis testing usually done

2.3 Analytical study designs

Analytical studies focus on the individual within the population. Main objectives of epidemiology are:

- Testing of hypothesis formulated from descriptive epidemiology
- Judging whether a particular exposure causes disease
- To identify the source, vehicle or risk factors for the disease
- To inform public health authority
- Research, training and learning

Classification

1. Cross sectional
 - a. Descriptive cross sectional study
 - b. Analytical Cross sectional study
2. Ecological
3. Longitudinal
 - a. Case control
 - b. Cohort

2.3.1 Cross sectional studies

Cross sectional study is the simple form of an observational study. It is based on a single examination of a cross section of population at one point in time. It is also known as prevalence study because it measures the prevalence of health problems.

In cross sectional study, the measurements of exposure and outcome are made at the same time. The key questions asked in this study is whether the exposure followed outcome or vice versa. Cross sectional studies are relatively easy and economical to conduct and are useful for investigating exposures that are fixed characteristics of individuals such as ethnicity, socio-economic status and blood group.

A cross sectional study may be classified into two type namely descriptive cross sectional and analytical cross sectional study.

Descriptive cross sectional study:

It is also known as prevalence study. It describes the current existing situation or events at a particular point in time.

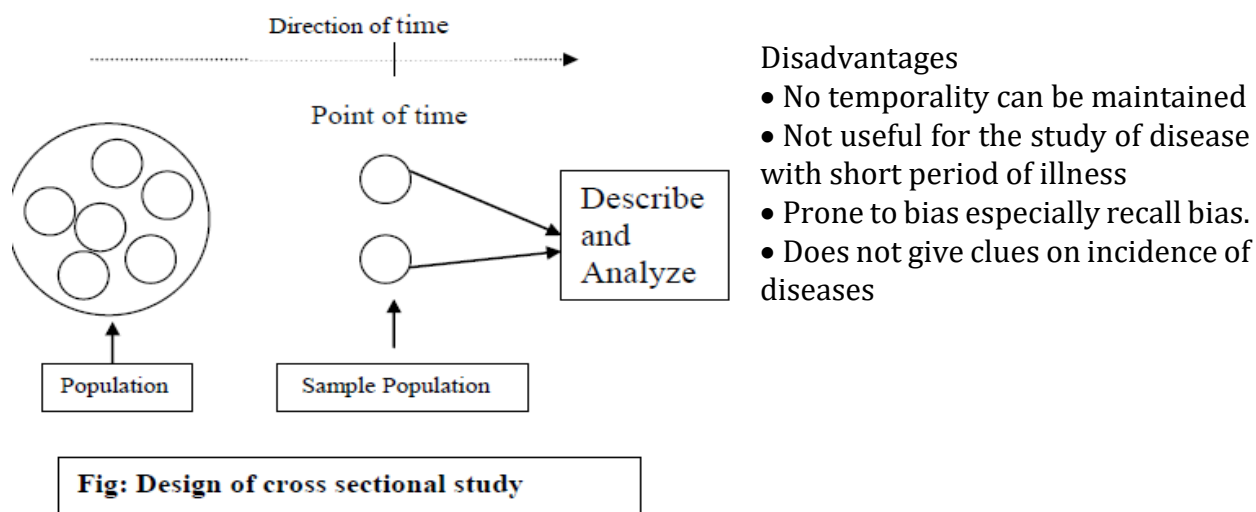
Analytical Cross Sectional Study:

In analytical cross sectional study, the data collection on both exposure and outcome takes simultaneously. When the prevalence of outcome is compared with the exposed and non exposed group, it is possible to identify the potential exposures on particular time. Analytical cross sectional studies are useful for defining the health needs of a population at a particular point in time and for investigating common exposures and common outcomes.

Advantages of Cross sectional studies

- Easy to perform

- Useful to measure current situation of health and disease problem
- Helps to measure the magnitude of health problem at one time
- identifies the risk factors or etiology of disease



2.3.2 Ecological study

Ecological or correctional studies also frequently initiate the epidemiological process. In ecological study, the unit of analysis is population or groups of people rather than individuals. The relationship of two variables may be studied by comparing population in different countries at the same time or same population in one country at different times. Example in _A|| country a relationship was demonstrated between average fat consumption and prevalence of obesity and coronary heart disease.

Although it is simple to conduct and thus attractive, ecological studies are often difficult to interpret since it is seldom possible to examine directly the various potential explanations for findings. These studies rely on the data collected for other purpose. Since the unit of analysis is population, individual link between exposure and effect can not be made. An ecological fallacy or bias results if inappropriate conclusions are drawn on the basis of ecological data.

An ecological fallacy is the bias that may occur because an association observed between variables on an aggregate level does not necessarily represent the association that exists at individual level. The association that has been exhibited in one place may not exist in other. Thus, these are also weak studies.

2.3.3 Longitudinal studies

Longitudinal studies are also similar to the cross sectional study where observations are repeated in same population over a prolong period of time by means of follow up examination. Cross sectional studies can be equated with photograph while longitudinal studies refer to a cine film. Longitudinal studies are helpful to:

- Study the natural history of disease and its future outcome
- Identify the risk factors of disease

- Find out the incidence of disease/health problem
- To identify the association between exposure and outcomes

These studies provide valuable information regarding cause and effect that can not be provided of former one. Explanation of longitudinal study designs is as follows

2.3.3.1 Case control study

Case controls studies are also called retrospective studies because the investigator looks backward from the disease to a possible causes. These are the first attempts to test causal hypothesis. These are relatively simple and economical to carry out and used to investigate the cause of disease especially rare diseases

Case control study has three features

- Both exposure and outcome has occurred before the start of study
- The study proceeds from backward from effects to cause
- Uses a comparison group to support or refute an inference

A case control study begins with the selection of cases and controls, which should represent all the cases from a specified population. The most difficult task is to select controls so as to sample the exposure prevalence in the population that generated the cases. The controls should represent people who would have been designated study cases if they had developed the disease.

To examine possible relation of an exposure to a certain disease, we identify a group of individuals with particular disease (called cases) and for purpose of comparison, a group of people without the disease (called control). We determine that proportions of the cases were exposed and what proportions were not exposed. We also determine what proportions of the controls were exposed and what proportions were not.

Thus, in a case control study, if there is an association of an exposure with a disease, the prevalence of history of exposure should be higher in persons who have the disease (cases) than in those who do not (controls).

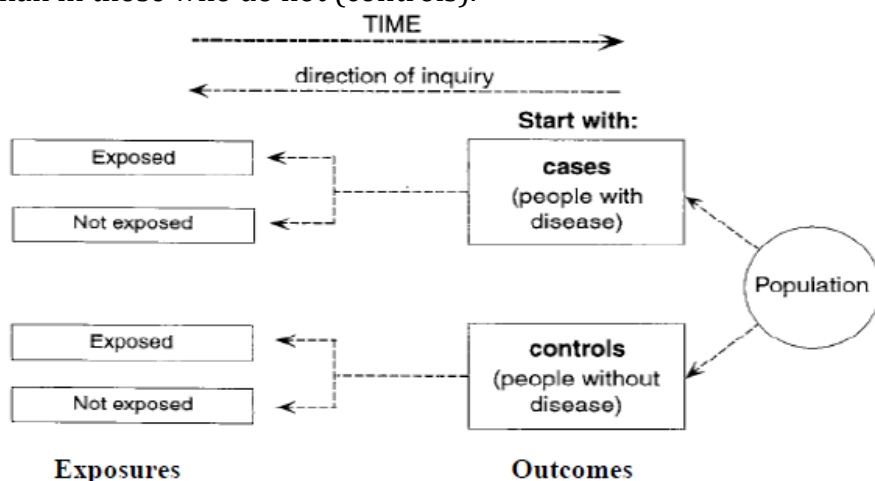


Fig: Design of a case control study

Let us consider the basic design of case control in 2x2 tables.

Table presents a hypothetical schema of how a case control study is conducted. There are **a** cases that were exposed and **c** cases that were not exposed. Similarly there are **b** controls that were exposed and **d** controls that were not exposed. Thus the total number of cases is $a+c$ and the total number of controls is $b+d$. If exposure is associated with disease, would expect the proportion of the cases who were exposed or $a/a+c$, to be greater than the proportion of the controls who were exposed or $b/b+d$.

Design of a case control studies (2x2)

	Cases (with disease)	Controls (with out disease)
Exposed	A	b
Not exposed	C	d

Basic steps/ procedures of case control study**1. Selection of the cases and controls**

The first step of this study is to select the cases and controls.

- Selection of cases**

The first step of this study is to select the cases and these can be easily found out than controls from the different sources. First of all, cases must be defined on the basis of two broad criteria.

Diagnostic criteria: The diagnostic criteria include case definition based on the different approaches such as clinical, operational, laboratory and any others. Once the diagnostic criteria set out, no alteration is made during the whole study period.

Eligibility criteria: The eligibility of the cases should be defined based on different factors such as age, sex, location or any other factors that must be designed before the study.

Source of cases

Sources of the cases should be designed before the starts of the study. The source of cases may be a variety.

- The source may be the hospital (s) patients, patient in physician's practice (clinic patients), and
- It can be selected from the community or entire population or randomly selected.
- The other sources of the cases may be drawn from the general population for a study period by general method or disease registry or vital statistics records. Cases must be representative of general population.

- Selection of controls**

The controls must be free from disease. This should be identified before the study. Generally the control group should be alike with the cases except disease outcome or disease interest.

Source of controls

The control group must be selected from the non-hospital sources i.e. from the community. It may be from the hospital source or admitted in hospital for disease other than that of disease of interest. Therefore the control group is selected from the neighborhood (neighbor control), relative controls, friend controls and general population.

The control obtained in this fashion may be likely to be similar to the case in age and in many other demographic and social characteristics.

2. Matching

Matching is defined as the process of selecting the controls so that they are similar to the cases in certain characteristics, such as age, race, sex, socioeconomic status, education, religion and occupation. It is the process of making comparison between the cases and controls, which are known to be influence the outcome of disease. The matching may be of two types.

i. Individual matching

This is also called paired matching. In this approach, each of cases selected for the study, a control is selected who is similar to the case in terms of specific variable or variables of concern. For instance, if first case is enrolled in our study is 40-year-old women; we will seek a 40-year female for control. If the second case is 20-year-old man from Aryan, we will select 20-year-old man from Aryan. This type of control selection is yields matched case control pairs, that is, each case is individually matched to a control.

ii. Group matching (frequency matching) It consists of selecting the controls in such a manner that the proportion of controls with a certain characteristics is identical to the proportion of cases with the same characteristics. Thus, if 25% of the cases are married, the controls will be selected so that 25% of that group is married.

Problems with matching Practical problem is difficult to identify appropriate matching. To minimize the problem, more variables are not taken for matching.

3. Measurement of exposure

Definitions and criteria about exposure are just as important as those used to define cases and controls. it is important to recognize that when case control studies are being used to test associations, the questions of bias must be ruled out. The exposure status can be observed through several techniques such as interview, hospital records, employment records, surveillance reports, disease registers etc.

	Cases (With disease)	Controls (With out disease)	Total
Exposed	a	b	a+b
Not exposed	c	d	c+d
Total	a+b	b+d	a+b+c+d
Exposure rate	a/a+c	b/b+d	

4. Analysis and interpretation

The final step of case control study is to analyze the data and interpret the results. It is used to find out:

- The exposure rates among cases and controls to suspected factors
- Estimation of disease risk associated with exposure (Odds ratio)

Using the information of above table, both exposure rates as well as odds ratio is calculated as follows

a. Exposure rates

Exposure rates among cases

$$= \frac{\text{Total no of exposed cases}}{\text{Total no. of cases}}$$

Thus, exposure rate among cases = $a/a+c$

Similarly, Exposure rates among controls

$$= \frac{\text{Total no. of exposed controls}}{\text{Total no. of controls}}$$

Thus, exposure rates among controls = $b/b+d$

b. Estimation of Risk associated with exposure

The association of an exposure and a disease is measured in case control study by calculating odds ratio which is the ratio of the odds of exposure among the cases to the odds in favor of exposure among the controls. This determines the risk associated with particular exposure. The derivation of odds ratio is based on following assumptions

1. disease being investigated must be relatively rare
2. the cases and control must be representative of diseased and disease free population

Odds ratio is given by:

OR = Odds of exposure among cases /odds of exposure among controls

Here,

$$\text{Odds of exposure among cases} = \frac{\text{Cases exposed to potential causal factor}}{\text{Cases not exposed to the particular factor}}$$

Thus from above table, Odds of exposure among cases = a/c

Similarly, odds of exposure among controls

Controls exposed to potential causal factor /controls not exposed to such factors

i.e. b/d

So, odds ratio

$$= \text{Odds of exposure among cases /odds of exposure among controls}$$

$$\frac{a/c}{b/d}$$

$$\text{Odds ratio} = ad/bc$$

This is also known as cross product ratio

	Cases (With disease)	Controls (With out disease)
Exposed	a	b
Not exposed	c	d

Example We are conducting a case control study of the relationship of smoking to coronary heart disease (CHD). We start with 200 people with CHD (cases) and compare them to 400 people without CHD (controls). If there is a relationship between smoking and CHD, we would anticipate that a greater proportion of the CHD cases than the control would have been smokers (exposed). We find that of the 200 CHD cases, 112 were smokers and 88 were non-smokers. Of the 400 controls, 176 were smokers and 224 were non- smokers. **Thus 56% of the CHD cases were smokers compared to 44% of the controls.** This calculation is only a first step. Further calculations can be done to determine whether or not there is an association of the exposure with disease.

Illustrative examples

	Cases (With disease-CHD)	Controls (With out disease)
Expose (Smoker)	112 (a)	176 (b)
Not Exposure (Non smoker)	88 (c)	224 (d)
	200 (a+c)	400 (b+d)
		176/400=44% b/(b+d)

Solution

- Exposure rates among cases
= $112/200 \times 100$
= 56 %
- Exposure rates among controls
= $176/400 \times 100$
= 44%

$$\text{OR} = 112 \times 224 / 88 \times 176$$

$$= 1.61$$

Interpretation This indicates that the cases were 1.61 times more likely than the controls have taken smoke.

Advantages:

- Quick, inexpensive, small sample size needed.
- Good for rare disease.
- Good for disease with long latent period.
- Can estimate multiple risk factors for a disease.

Disadvantages:

- Prone to bias.
- Can't measure incidence rate.
- Difficult to determine temporality between exposure & disease.
- Not good for rare exposure.

Bias in case control study

Bias is any systematic error in the determination of association between the exposure and outcome. The common errors observed beings are

1. bias due to confounding
2. Recall Bias/memory bias
3. Selection Bias
4. Berksonion Bias: the Bias arises due to different rates of admission to hospitals(hospital cases and control)
5. Interviewer bias

2.3.3.2 Cohort Study

A well designed cohort study is considered the most reliable means of showing an association between or suspected risk factor and subsequent disease, because it eliminates many of the experimental models of physical sciences. It is important analytical study to refute or prove the existence of an association between suspected cause and disease.

Features of cohort study

1. The prospect (future effect) of a suspected cause is studied. – Prospective study
2. This study proceeds forward from cause to effect. – Forward looking study
3. The study groups are observed over a period of time. – Longitudinal study
4. The frequency of diseases (incidence) can be determined. – Incidence study
5. The study is done among the cohorts. – Cohort study

What is Cohort? Cohort is defined as a group of people who share a common characteristic or experience within a defined time period. **Example;** Birth cohort, Occupation cohort, Exposure cohort, Marriage cohort, etc.

Definition of cohort study A type of analytic study which groups of individuals defined on the basis of their exposure status are followed to assess the occurrence of disease

Indications for Cohort Study

1. When there is good evidence of an association between exposure and disease.
2. When exposure is rare but incidence of the disease is high among the exposed.
3. When follow up of study population is easy.
4. When the cohort is stable.

5. When the study population is accessible and cooperative.
6. When there is sufficient fund to carry out the study.

Types of Cohort Studies These types of cohort studies have been distinguished on the basis of the time of occurrence of disease in relation to time at which the investigation is illustrated and continued.

1. Prospective Cohort Studies:

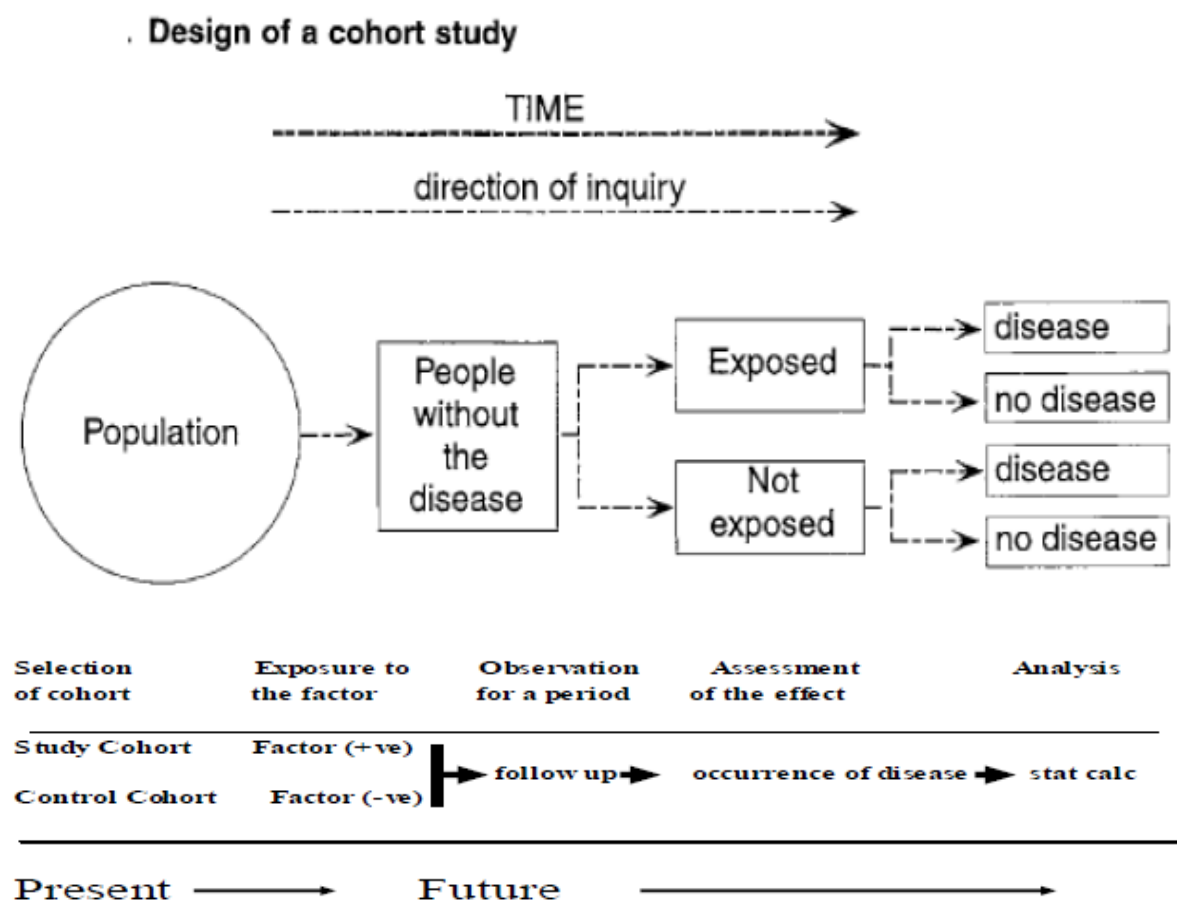
Concurrent cohort: A prospective cohort study is one in which the outcome has not yet occurred at the time the investigation begins in the present and continue into future.

2. Retrospective Cohort Study:

A retrospective or historical cohort study is one in which the outcomes have all occurred before the start of investigation, the investigator goes back in time.

3. Retro-Prospective Study

In this type of study, both the retrospective and Prospective elements are combined. The cohort is identified from past records and it's assessed of data for outcome. The same cohort is followed up prospectively into future for further assessment of outcomes.



Steps of carrying out a cohorts study

1. Selection of Study subjects.
2. Obtaining data on exposures.
3. Selection of comparison groups.
4. Follow-up
5. Analysis and interpretation

1. Selection of study elements:

The subjects are usually assembled either from general population that can be readily studied. Study subjects are selected from

- General Population
- Special group.

2. Obtaining data on exposure

Information about exposure may be obtained directly from the cohort members through personal interview or mailed question mare

- a. Review of records: Certain kinds of information can be obtained only from medical records.
- b. Medical examination on special tests: Some types of information can be obtained only by medical examination.
- c. Environmental Surveys: This is the best source for obtaining information on exposure.

3. Selection of Comparison Groups:

There are many ways of assembling comparison groups.

- a. Internal Comparison: Comparison within groups.
- b. External Comparison: When information or degree of exposure is necessary to put up an external control.

4. Follow up

One of the problems in cohort studies is the regular follow-up of all the Participants. Therefore, at the start of the study the methods should be devised depending upon the outcome to be determined.

5. Analysis and interpretation

The data analysis can be done in terms of

- incidence rates of outcomes among exposed and non exposed
- estimation of risk associated with particular exposure/outcome

Incidence rates of outcomes among exposed and non – exposed In cohort study, we can determine incidence rates directly in those exposed and non- exposed. Let us consider a symbolic example of in 2x2 tables.

	With disease	With out disease	Total
Exposed	a	b	a+b
Not exposed	c	d	c+d
Total	a+b	b+d	a+b+c+d
Exposure rate	a/a+c	b/b+d	

Rate of incidence of disease among exposed people

$$a/a+b$$

Rate of incidence of disease among non- exposed people

$$c/c+d$$

Estimation of risk

The risk associated with specific exposure to the particular outcome can be determined by following statistical calculations

a. Relative risk

Relative risk is the ratio of incidence of disease among exposed and non - exposed. It is also called risk ratio. It is one of the methods to exhibit the statistical association between exposure and outcome. It is denoted by RR and given by.

RR = Incidence of disease among exposed groups/ Incidence among non- exposed

$$\frac{a/a+b}{c/c+d}$$

Interpretation

- A relative risk of 1 means no association between suspected cause and exhibited outcome
- A relative risk of greater than 1 means there is strong association between cause and disease.
- A relative risk of less than 1 suggests that the factor is not causal but ct as protective.

Estimation of relative risk is important in the measurement of strength of association between suspected cause and effect.

Attributable risk

Attributable risk or risk difference or excess ratio is the difference in the rates of disease between the incidence rates of disease among exposed and incidence rates among non-exposed. AR = Incidence among exposed – incidence among non-exposed.

It can also be calculated as

AR = incidence of disease among exposed – incidence of disease among non exposed

Incidence of disease among exposed group

This is also called relative attribution.

Interpretation

Attributable risk indicates to what extent the disease under study can be attributed to exposure

Strengths/Advantages

- Correct temporal relationship between exposure and the outcome of the disease
- Disease rates can be calculated directly
- Allows the evaluation of multiple effects from a single exposure
- Generally involved good information on exposure

Limitations/Disadvantage

- Large number of subjects are required if the outcome of interest is rare
- Requires long follow-up, if latency period is long
- Expensive
- Investigator should consider several issue: availability of study groups, records, and feasibility of complete follow-up of subjects

2.4 Experimental Studies

Intervention or experimentation involves attempting to change a variable in one or more groups of people. The effects of intervention are measured by comparing the outcome in experimental group with that in a control group. Since the interventions are strictly determined by the protocol, ethical considerations are of paramount importance on design of these studies. Experimental studies can be classified as

- a. True experiment
- b. Quasi - experimental studies: these are the experiment in which one or more experimental component is lacking.

True experimental studies are further classified into three types 1. Clinical trial: controlled and uncontrolled/randomized and non randomized trials 2. Community trial 3. Field trial

2.4.1 True Experimental studies

These are the experiments done on the scientific basis. Such studies take references of comparison group to support or refute the outcome with the help of established evidences. A brief explanation of each type of experimental study is as follows:

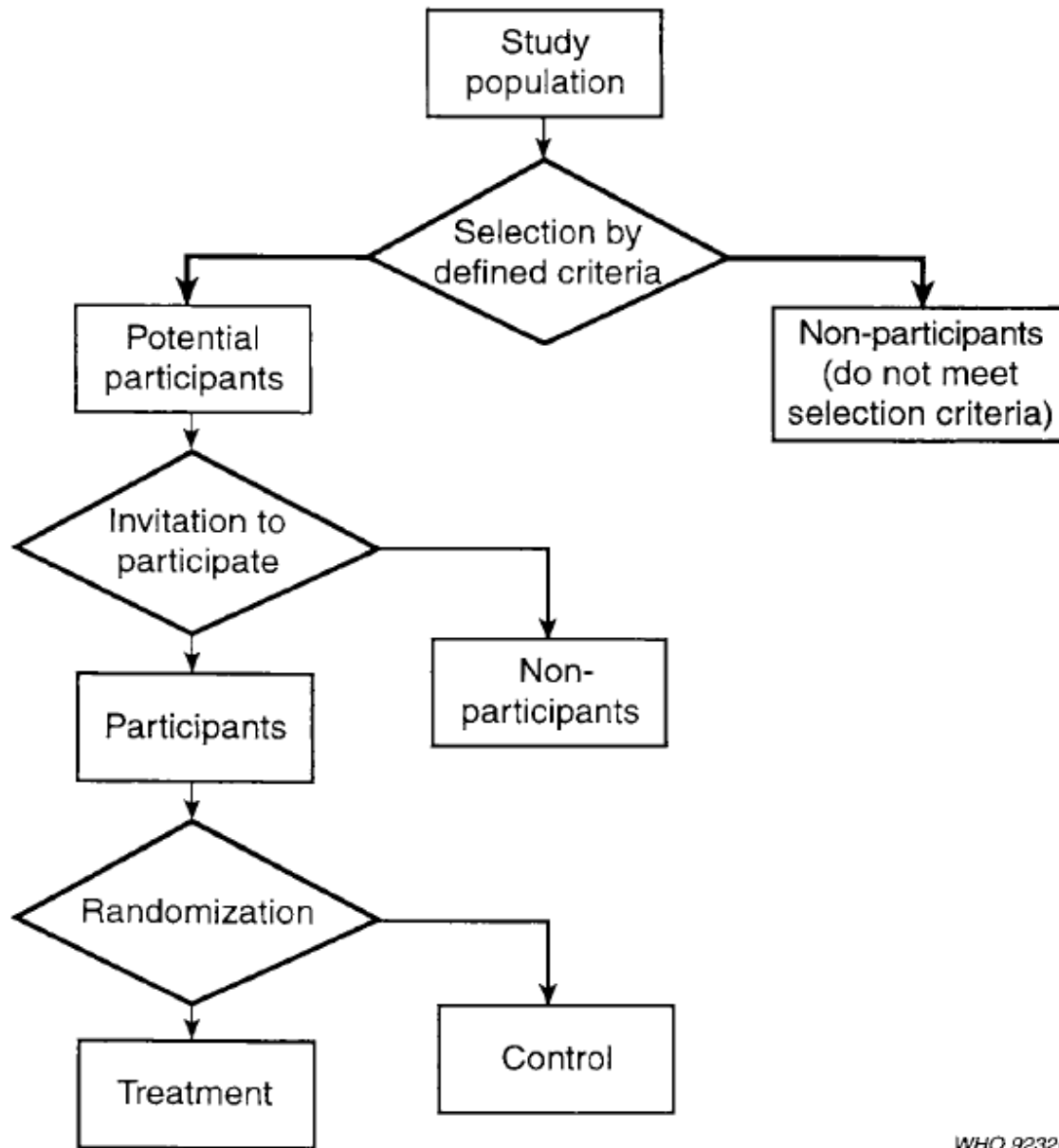
1. Clinical Trial

It may be controlled as laboratory experiment (controlled trial) or non controlled i.e. Study is done in natural setting (uncontrolled trial). It may be done through either random allocation of participants (randomized trial) or purposive division (non randomized trial). The best study is randomized controlled trial which preserves the scientific values, allows statistical calculation and confounding never occurs.

1. a. Randomized clinical /controlled trial It is one of the important scientific method to evaluate the effectiveness of interventions or preventive or therapeutics (treatments). A randomized controlled trial is an epidemiological experiment to study a new preventive or therapeutic regimen. Subjects in a population are randomly allocated to groups, usually

called treatment and control groups and the results are assessed by comparing the outcome in the two or more groups.

Design of Randomized Clinical Trial



WHO 92326

Steps in carrying out Randomized controlled trial The basic steps in conducting a randomized controlled trial include following sequential activities.

1. Drawing up a protocol
2. Selecting reference and experiment population
3. Randomization

4. Manipulation or intervention
5. Follow up
6. Assessment of outcome

Description of each step is as follows:

1. Drawing up a protocol

One of the essential features of RCT is that it is performed under strict protocol. The protocol specifies the aim and objectives of the study; criteria for selection of study and control groups, sample size, procedure for allocation of subjects and treatments to be applied.

2. Selecting reference and experiment population

a. Reference population

It is the population to which the findings of trial, if found successful are expected to be applicable. The reference population may comprise the population of a whole city or a population of school children and so on according to the nature of the study.

b. Experimental or study population

It is the actual population that participates in experimental study. Ideally, it should be randomly chosen from reference population, so that it has similar characteristics as the reference population. Once the experimental population has been defined, its members are invited to participate in the study. The participants or volunteers must fulfill the following criteria

- Must give informed consent
- Must be representative of population
- Qualify or eligibility for trial

3. Randomization

It is a statistical procedure by which the participants are allocated into groups usually called study and control or placebo group to receive or not to receive an experimental or preventive or therapeutic procedure. Randomization is an attempt to eliminate bias or error. It is heart of the controlled trial. It will give the greatest comparability so that like can be compared with like. Randomization is done only after the participants have entered into the study.

4. Manipulation or intervention The manipulation creates an independent variable whose effect is then determined by measurement of final outcome which constitutes the dependent variable.

5. Follow up This implies examination of the experimental and control group subjects at defined interval of time in standard manner, with equal intensity. The follow up may be short or may require many years depending upon the study undertaken.

6. Assessment of Outcome The final step is to assess the outcome of trial in terms of positive results; that is benefits of experiment or effectiveness of intervention. Negative results is observed in terms of increase in severity, frequency and side effects and complications. The incidence of results rigorously compared in between experiment and placebo group.

Blinding To reduce the possible error while attempting epidemiological studies, a process in which one or more person concerned with experiment are unknown or blind about the status of study participants is called blinding. Blinding can be classified into three types

- a. Single blinding: the trial is so planned that the participants are not aware whether they belongs to experiment or placebo group.
- b. Double blinding: The trial is so planned that neither the doctor nor participants are aware of group allocation and treatment received.
- c. Triple Blinding: The participants, investigator and the person analyzing the data are all blind. Ideally of course, triple blinding should be used in RCT.

Example of RCT

A trial using rice based or glucose based oral re-hydration solution involves 342 Patients with acute watery diarrhea during an epidemic of cholera in Bangladesh in 1983(Molla et al 1985). The patients were randomly assigned to treatment with either glucose based or rice based oral re-hydration solution. The study showed that the glucose component of oral re-hydration solution could be replaced by rice water with improved results, as indicated by decrease in mean stool output and intake of solution.

2. Community Trial

In this form of experiment, the treatment groups are communities rather than individuals. This is particularly appropriate for diseases that have their origins in social conditions, which in turn can most easily be influenced by intervention directed at group behavior as well as at individuals. Cardiovascular disease is a good example of condition appropriate for community trials, several of which have been completed. A limitation of such studies is that only a small number of communities can be included and random allocation of communities is not practicable. Other methods are required to ensure that any differences found at the end of study can be attributed to the intervention rather than no inherent differences between communities. Furthermore, it is difficult to isolate the communities where interventions are taking place from general social changes that may be occurring. Consequently this type of study may underestimate the effect of intervention.

3. Field trials

Field trials, in contrast to clinical trial, involve people who are disease free but presumed to be at risk; data collection takes place in the field. Usually among non -institutionalized people in the general population. Since the subjects are disease free and the purpose is to prevent the occurrence of diseases that may occur with relatively low frequency. For example, one of the largest field trials ever undertaken was that of the Salk vaccine for the prevention of poliomyelitis, which involved over million children. The field trial method can be used to evaluate interventions aimed at reducing exposure without necessarily measuring the occurrence of health effects. Such intervention studies can often be carried out on a small scale at low cost.

2.4.2 Quasi experimental studies

These are weak experimental designs in which at least one components of true experiment is absent. For example either there is lack of comparison group or randomization or others.

Example of quasi experiment is imparting health education to community people and its evaluation without reference to comparison group.

2.5 Strategies of epidemiology

Strategy is an attempt with the help and evaluation provided by history to draw guideline that favors and applies there will be the perspective. A guideline may favor success or decreases the likelihood of failure. Epidemiology focuses on defining and describing the events by time, place and person and formulation of ideas to establish new facts. The major concern of epidemiology is to identify the determinants of health or health problems by establishing cause and effect relations i.e. causal association. These may be either descriptive or analytical or experimental strategies.

Major strategies applied in epidemiology are

1. Description of distribution of disease
2. Forming hypothesis
3. Stating hypothesis
4. Testing hypothesis

1. Description of distribution of disease

The initial approach in epidemiology is to describe the distribution of disease in relation to the specific characteristics such as time, place and person. It is evident that the variables listed are not necessarily those most likely to have causal association with disease being investigated.

2. Forming hypothesis The success or failure of research frequently depends upon the soundness of the hypothesis. New and convincing hypothesis can be one of the most powerful forces influencing the direction of future researches. The process of formulation of hypothesis can be classified into four types

- A. Method of difference
- B. Method of agreement
- C. method of concomitant variation
- D. Analogy

A. Method of difference : If the frequency of a disease is markedly different under two separate circumstances and some factors can be identified in some circumstances that is absent in other, this factor or its absence may be identified as a cause of disease. It is useful in determination of hypothesis related to deficiency disorders.

B. Method of agreement : If a factor is common to a number of different circumstances that are associated with the presence of the disease, this factor may be the cause of disease. For example: cervical cancer is associated with multiple sexual partners and low socioeconomic status. A common factor may be a virus transmitted through venereal contact.

C. Method of Concomitant Variation : The method of concomitant variation merges interceptibility into the two previously described methods because in many situations it is not a matter of a factor being present or absent but of its being present in a greater or lesser

degree. An excellent example is: An attempt has been made to relate the various dietary constituents to the risk of coronary heart disease in different parts of world.

D. Analogy : The distribution of disease may be sufficiently similar to that of some other disease that has been more completely and successfully investigated as to suggest that certain cases may be common to both.

Thus we tend to associate a disease occurring in parts of the world during summer months with vector transmission.

E. Hypothesis that arises from data : Specific hypothesis come to attention not from any rational consideration of distribution disease and its possible determinants.

3. Statement of Hypothesis Ideally an epidemiological hypothesis should specify the followings;

1. The characteristics of persons to whom the hypothesis applies
2. The cause being considered — environmental factors
3. The expected effect ---- the disease
4. The exposure – response relationship
5. The time response relationship

4. Testing of hypothesis Hypothesis can be tested by carefully designed analytical studies or even more with experimental studies. A little evidence can be provided by ecological studies and statistical relationship can be seen from case control as well as cohort studies.

Some important concepts used in epidemiological tests

Validity

Validity is an expression of degree to which a test is capable of measuring what is intended to measure. A study is valid if its results correspond to the truth i.e. there should be no systematic error and the random error should be very small.

Validity can be classified into two types:

1. Internal validity
2. External validity

1. Internal Validity:

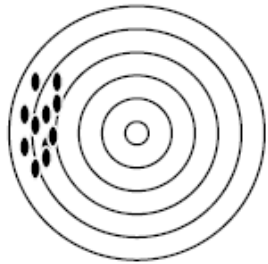
It is the degree to which the results of observation are correct for the particular group of people being studied. Example: management of blood hemoglobin. The analysis of blood in different laboratory may produce different results but the evaluation of association with anemia as measured by one laboratory may be still internally valid. Internal validity may be threatened by the sources of systematic error.

2. External validity:

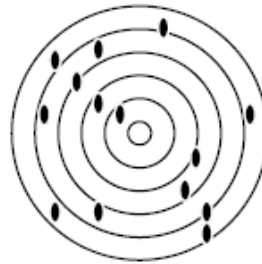
External validity or generalizability is the extent to which the results of a study apply to people not in it. Internal validity is necessary for but does not guarantee external validity.

External validity requires external quality control of the measurements and judgments about the degree to which results of the study can be extrapolated.

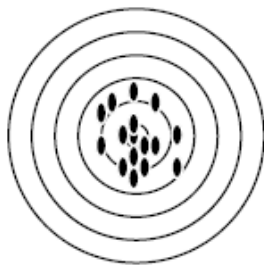
Reliability It is defined as the degree of stability exhibited when measurement is repeated under identical conditions. Reliability refers to the degree to which the results obtained by a measurement can be replicated. Lack of reliability may arise from divergences between observers or instruments of measurements or instability of the attribute being measured.



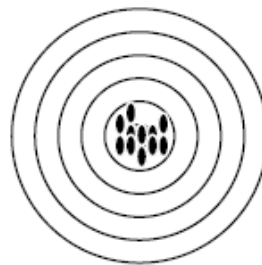
Score are highly reliable, but not valid



Score are neither reliable nor valid



Reliability has improved but is still low; scores are somewhat valid.



Scores are both reliable and valid.

Validity and reliability can be illustrated from given figure also.

CHAPTER 3

Infectious Disease Epidemiology

3.1. Terminology Used In Epidemiology

Infection

- The entry and development or multiplication of an infectious agent in the body of man or animals
- It also implies that the body responds in some way to defend itself against the invader, either in the form of an immune response or disease
- An infection does not always cause illness

Contamination

- The presence of an infectious agent on a body surface; also on or in clothes, beddings, toys, surgical instruments or dressings, or other inanimate articles or substances including water, milk and food

Infestation

- For persons or animals the lodgement, development and reproduction of arthropods on the surface of body or in clothing eg; lice, itch mite
- Term also use to describe invasion of the gut by parasitic worms eg. Ascariasis
- Infested articles or premises are those which harbor or give shelter to animal forms, especially arthropods and rodents

Host

- A person or other animal, including birds and arthropods, that affords subsistence or lodgement to an infectious agent under natural conditions
- Obligate host: means the only host eg, man in measles and typhoid fever
- Primary/definitive host: where parasite attains maturity or passes its sexual stage
- Secondary/intermediate host: Those in which the parasite is in a larval or a sexual state

Infectious disease

- A clinically manifest disease of man or animals resulting from an infection

Communicable disease

- An illness due to a specific infectious agent or its toxic products capable of being directly or indirectly transmitted from man to man, animal to animal , or from environment to man or animal

Epidemic

- The unusual occurrence of disease or health related events (traffic accident) clearly in excess of usual occurrence in a community or region

Endemic

- It refers to the constant presence of a disease or infectious agent within a given geographic area or population group, without importance from outside
- The usual or expected frequency of the disease within such area or population group
- An endemic disease when conditions are favorable may burst in to an epidemic (eg; Hepatitis A, typhoid fever)

Sporadic

- The word sporadic means scattered about
- The cases occur irregularly, haphazardly from time to time and generally infrequently
- Cases are so few and separated widely in space and time e.g. polio, tetanus, herpes zoster and meningococcal meningitis
- A sporadic disease may be the starting point of an epidemic when conditions are favorable of its spread

Exotic

- Disease which are imported in to a country in which they do not otherwise occur, as for e.g. rabies in UK

Zoonosis

- An infection or infectious disease transmissible under natural conditions from vertebrate animals to man e.g. Rabies, plague, bovin tuberculosis, Brucellosis

Epizootic

- An outbreak of disease in an animal population

Enzootic

- An endemic occurring in animal population

Nosocomial infection

- Is an infection originating in a patient while in a hospital or other health care facility

Opportunistic infection

- This is infection by an organism that takes the opportunity provided by a defect in host defence to infect the host and hence cause disease
e.g. tuberculosis infection in the condition of HIV

Iatrogenic infection

- Any unwanted or an adverse consequences of a preventive, diagnostic or therapeutic regimen or procedure that causes impairment, handicap, disability or death resulting from a physician's professional activity or from the professional activity of other health persons

Surveillance

- The continuous scrutiny of the factors that determine the occurrence and distribution of disease and other conditions of ill health
- Surveillance is essential for effective control and prevention, and includes the collection, analysis, interpretation and distribution of relevant data for action
- Thus we have epidemiological surveillance, nutritional surveillance, demographic surveillance, serological surveillance
- The main purpose of surveillance is to detect changes in trend or distribution in order to initiate investigative or control measures

Agent

- A factor, such as a microorganism, physical, chemical substance, or form of radiation, whose presence, excessive presence, or relative absence is essential for the occurrence of a disease

Analytic epidemiology

- The aspect of epidemiology concerned with the search for health-related causes and effects. Uses comparison groups, which provide baseline data, to quantify the association between exposures and outcomes

Applied epidemiology

- The application or practice of epidemiology to address public health issues

Association

- Statistical relationship between two or more events, characteristics, or other variables

Biologic transmission

- The indirect vector-borne transmission of an infectious agent in which the agent undergoes biologic changes within the vector before being transmitted to a new host

Chain of infection

- A process that begins when an agent leaves its host through a portal of exit, and is conveyed by some mode of transmission, then enters through an appropriate portal of entry to infect a susceptible host

Incubation period

- The period of time between the entry of an infectious agent and appearance of first sign or symptoms of disease. The infectious agent undergoes multiplication during the incubation period. Sufficient density of disease agent is needed to overtake the disease.

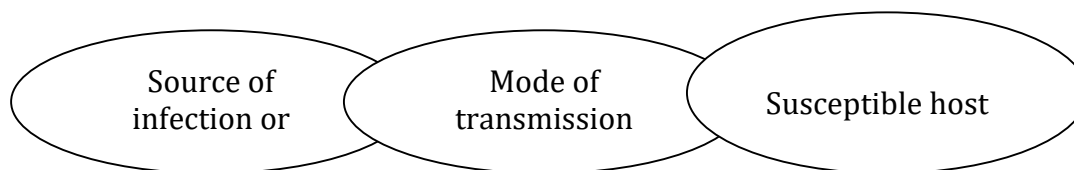
Communicable period

- The time during which an infectious agent may be transferred directly or indirectly from an infected person to another
- The communicable period varies in different diseases. Some diseases are more communicable during incubation period while others are more communicable during actual illness. The early diagnosis and treatment can reduce the communicability of disease.

3.2. Mode of Disease Transmission

Dynamics of disease transmission

Communicable diseases are transmitted from the reservoir or source of infection to susceptible host. Basically there are three links in the chain of transmission which is, the reservoir, mode of transmission and the susceptible host.



In the absence of these factors, the communicable disease will not occur.

1. Source or Reservoir

The starting point for the occurrence of a communicable disease is the existence of a reservoir or source infection. The source of infection is defined as the person, animal, object or substance from which an infectious agent passes to the host.

The reservoir is defined as “any person, animal, arthropod, plant, soil or substance in which an infectious agent lives and multiplies, on which it depends primarily for survival, and where it reproduces itself in such manner that it can be transmitted to susceptible host. In short, the reservoir is the natural habitat in which the organism develop and reproduce.

E.g. In hookworm infection, the reservoir is man and the source of infection is contaminated soil with infective larvae. In tetanus infection the reservoir and source are the same i.e. soil. In typhoid infection the reservoir is man, and the source of infection is faeces of patients or contaminated food, milk or water.

Types of reservoir

1. Human reservoir
2. Animal reservoir
3. Reservoir in non-living thing

1. Human reservoir:

The most important source of reservoir of infection for humans is man himself. He may be case or carrier.

a. Cases

A case is defined as “a person in a population or study group identified as having the particular diseases, health disorder or condition under investigation”. A variety of criteria (eg, clinical, biochemical, laboratory) may be used to identify cases. Broadly the presence of infection in a host may be clinical, subclinical or latent

i. Clinical cases:

The clinical cases may be mild or moderate, severe or fatal depending up on the level of involvement. This is clinically diagnosed disease with having sign and symptoms.

ii. Sub-clinical cases:

The disease is not diagnosed clinically with sign and symptoms but can be detected by laboratory test. Eg. Hepatitis A and B

iii. Latent infection:

The infectious agent lies in dormant form with in the host without symptoms and often not presence in blood, tissue or body secretion of host. Eg, window period in HIV infection

b. Carriers

In some diseases, either due to inadequate treatment or immune response, the disease agent is not completely eliminated, leading to carrier state. A carrier is defined as “an infected person or animal that harbours a specific infectious agent in the absence of discernible clinical disease and serves as a potential source of infection for others”.

The elements in a carrier states are:

- Disease agent presence in the body of the host
- The absence of recognizable symptoms and signs of disease
- The shedding of the disease agent through the discharges or excretions thus acting as a source of infection for other persons

2. Animal reservoir

The source of infection may sometime be animals and birds. The diseases and infections which are transmissible to man from vertebrates are called zoonoses. There are over 100 zoonotic diseases which may be conveyed to man from animals and birds. Eg. Rabies, influenza etc.

3. Reservoir in non-living things:

Soil and inanimate matter can also act as reservoirs of infection. For eg: soil may harbor agents that cause tetanus, anthrax etc.

2. Mode of Transmission

Communicable diseases may be transmitted from the reservoir or source of infection to a susceptible individual in many different ways. The mode of transmission of disease is the route of infectious agent passage from infected person or animal to another. The following are the mode of transmission:

A. Direct transmission

- Direct contact
- Droplet infection
- Contact with soil
- Inoculation in to skin or mucosa
- Transplacental (vertical)

B. Indirect transmission

- Vehicle-borne
- Vector- borne

- iii. Air-borne
 - Droplet nuclei
 - Dust
- iv. Fomite- borne
- v. Unclean hands and fingers

A. Direct transmission

- i. Direct contact

Infection may be transmitted by direct contact from skin to skin, mucosa to mucosa, or mucosa to skin of the same or another person. This implies direct and essentially immediate transfer of infectious agents from the reservoir or source to susceptible individual, without an intermediate agency eg: skin to skin contact as by touching, kissing or sexual intercourse. Diseases transmitted by direct contact include STDs and AIDS, leprosy, skin and eye infection
- ii. Droplet infection

This is direct spray of droplet of saliva and naso-pharyngeal secretions during coughing, sneezing or speaking and spitting, in to the surrounding atmosphere. The droplets with infectious agents enter in to respiratory tract eg. TB, diphtheria, whooping cough etc. the droplet spread is increased in conditions of overcrowding and lack of ventilation
- iii. Contact with soil

The disease agent may be acquired by direct exposure of susceptible tissue to the disease agent in soil eg: hookworm larvae, tetanus etc.
- iv. Inoculation in to skin or mucosa:

The disease agent may be inoculated directly in to the skin or mucosa eg; rabies virus by dog bite, hepatitis B virus through contaminated needles and syringes etc.
- v. Transplacental (vertical)

Disease agent can be transmitted from pregnant mother to her fetus eg: HIV, TORCH agents, Hepatitis B etc

B. Indirect transmission

- i. Vehicle-borne:

Transmission by contaminated food, water, milk, fruits etc. the infectious agent may have multiplied or developed in the vehicle (eg. *S. aureus* in food) before being transmitted; or only passively transmitted in the vehicle eg. Hepatitis A virus in water. Disease transmitted by water and food chiefly infections of alimentary tract eg, acute diarrhoea, typhoid fever, cholera, polio, hepatitis A, food poisoning etc.

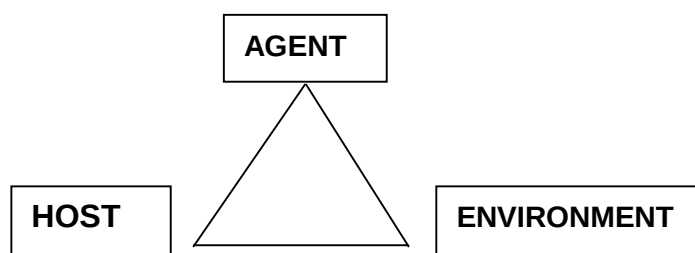
- ii. **Vector –borne:**
In infectious disease epidemiology, vector is defined as an arthropod or any living carrier that transports an infectious agent to a susceptible individual, with mechanical or biological process. Eg, malaria, kalazar
- iii. **Air borne:**
Some bacteria and virus remain in the air with dust and droplet nuclei. These infectious agents enter the human body during respiration. Eg: tuberculosis, measles, influenza etc.
- iv. **Fomite borne:**
Fomite are the inanimate articles or things eg: clothes, towels, handkerchief, pencils, books, toys, cups, spoons etc. disease transmitted by Fomites are diarrhoea, dysentery, hepatitis 'A', eye infection etc.
- v. **Unclean hands and fingers:**
Dirty hands fingers are the common source of infections. The disease which transmits by unclean hands and fingers are typhoid, dysentery, Hepatitis 'A', worms infestation, diarrhoea etc.

3. Susceptible host

Susceptible host means the host which is easily influenced for infection. The host became susceptible when the disease agent and environment are in favorable condition. It means there is potential chance of interacting the agent, host and environment.

3.3. Epidemiological triad

- The epidemiological triad is the interrelationship between disease agent host and environment.
- Only disease agent is not sufficient to produce disease
- The other factor relating to host and environment is equally important whether disease will occur or not in exposed host



DISEASE AGENT

- Disease agent is defined as any substance living or non-living or a force tangible or intangible the excess presence or relative lack of which may initiate a disease process

Classification of disease agent

- Biological agent: these are living agents eg: Bacteria, virus, fungi, protozoa

- Nutritional agent: protein, fats, CHO, Vitamin, minerals and water. Any excess or lack may produce disease
- Physical agent: heat, cold, pressure, radiation, electricity, noise may produce disease
- Chemical agent: Acid, alkali, metals, allergens, insect
- Social agent: poverty, smoking, drug abuse, unhealthy lifestyle, social isolation

HOST

- A person or animal including birds and arthropods that affords to alive and lodgement to an infectious agent under natural conditions
- In epidemiological terminology human host is consider “soil” and the disease agent is taken as “seed”
- The host factor play a major role to determine the outcome of disease in the exposed individual

Classification of host factor

- Demographic characteristics: eg. Age, sex, ethnicity
- Biological characteristics: eg. Genetic factors, biochemical level of blood, immunological factors
- Social and economic characteristics: eg, education, occupation, economic status, stress, housing condition
- Lifestyle factors: habits, exercise etc.

ENVIRONMENTAL FACTORS

- The environmental factor play vital role to produce disease or health problems. The low or greater health problems are depend on the sanitary condition where the people live. Mainly the 3 environmental factors are more or less responsible for creating health problems:
- Physical environment: nonliving things eg; air, water, housing, heat, light, noise, radiation, cold, geography, climate etc
- Biological environment: bacteria, virus, insects, rodents, animals, plants and human beings. These are constantly working for their survival and in this process some of them act as disease producing agent, reservoir of infection, and vector of disease
- Psychosocial environment: the psychosocial environment includes culture, values, customs, habits, belief, attitude, moral, religion, education, lifestyle, health service available etc
- An individual is constant interaction with all the above psychosocial environment and these environment directly or indirectly affect the health of individuals
- Poverty , loss of employment, urbanization, migration, exposure to stressful situation may create feeling of anxiety, depression, anger, frustration etc. which initiate disease process

3.4. Disease Prevention and Control

Infectious disease control and prevention is the breaking the chain of infection. The following are the main method of infectious disease control and prevention.

- A. Controlling the reservoirs
- B. Interruption of transmission
- C. Protecting the susceptible host

A. Controlling the reservoir

The most desirable control measure would be to eliminate the reservoir or source, if that could be possible. Elimination of the reservoir may be pretty easy with the animal reservoir, but is not possible in humans in whom the general measures of reservoir control comprises:-

1. Early diagnosis

1st step in the control of communicable disease is the rapid identification. This is a very much important activity for disease control. Frequently, laboratory procedures may be required to confirm the diagnosis.

2. Notification

Once an infectious disease has been detected or even suspected it should be notified to the local health authority whose responsibility is to control measures including the provision of medical care to patients. Notification is an important source of epidemiological information. It enables early detection of disease outbreaks, which permits immediate action to be taken by the health authority to control there spread. Example: measles outbreak can be controlled if only single case can be notified properly in time.

3. Isolation

Isolation is oldest communicable disease control measure. It is the separation of infected person from others during the period of communicability. The purpose of isolation is to protect the community by preventing transfer of infection from the reservoir to the possible susceptible hosts.

4. Treatment

Many communicable disease have been controlled by effective drugs. The objective of treatment is to kill the infections agent when it is still in the reservoir. Treatment reduces the communicability of disease, short the duration of illness and prevent development of secondary cases. In some disease e.g. syphilis, TB, and leprosy. Early diagnosis and treatment is of primarily importance in interrupting transmission. Treatment is also extended to carriers.

Treatment can take the form of individual treatment or mass treatment. In the latter category, all the people in the community are administrated the drugs whether they have the disease or not (e.g. trachoma)

5. Quarantine

Quarantine has been defined as the limitation of freedom of movement of such persons or domestic animals exposed to communicable disease for a period of time not longer than usual incubation period of the disease. The purpose of quarantine is to prevent effective contact with those people not so exposed.

B. INTERRUPTION OF TRANSMISSION

It is the breaking the chain of infection by changing some components of man environment such as:

- Clean drinking water
- Drinking safe water
- Hand washing
- Maintain food hygiene
- Control vector of disease
- Maintain clean environment
- Reduce pollution

C. PROTECTING THE SUSCEPTIBLE HOST

The susceptible host can be protected by following methods:

i. Active immunization:

One of effective way of controlling the spread of infection is to strengthen the host defences. It is one of the most powerful and cost-effective weapons of modern medicine. There are some infectious diseases whose control is solely based on active immunization. Active immunization produce antibodies for specific disease prevention eg:- BCG, DPT, Polio, Measles, Hepatitis etc.

ii. Passive immunization:

It is the transfer of antibodies to induce protection against disease. Passive immunization is a short-term expedient useful only when exposure to infection has just occurred. Eg:- Rabis vaccination, tetanus toxoid, Anti-snake venom etc.

iii. Combined passive and active immunization:

In some particular disease, the combined method is effective eg, hepatitis B
Hepatitis B vaccination + Hepatitis B immunoglobulin (in case of Hep. B exposure)

iv. Chemoprophylaxis:

Administration of drugs before or after exposure in certain factors or disease agent but sign and symptoms of disease has not been developed.

Eg. Plague : Tetracycline for contacts of plague

Bacterial conjunctivitis : Erythromycin ophthalmic ointment.

v. Non-specific measures:

Most of the non-specific measures to interrupt pathways of transmission are in general applicability. Improvement water supply, sanitation, nutrition education fall in to this category.

Non- specific measures will also include 'legislative measures' such as pollution in control. Another important non-specific measure is community involvement in disease surveillance, disease control and other public health activities.

CHAPTER 4

Clinical Epidemiology

4.1. Introduction of Clinical Epidemiology

Concept and definition

Clinical epidemiology was introduced by –John Paul in 1938. According to him “a new basic science for preventive medicine”. The shift in the focus of clinical epidemiology from community ecology to individual patients and groups of patients took place in the 1960s

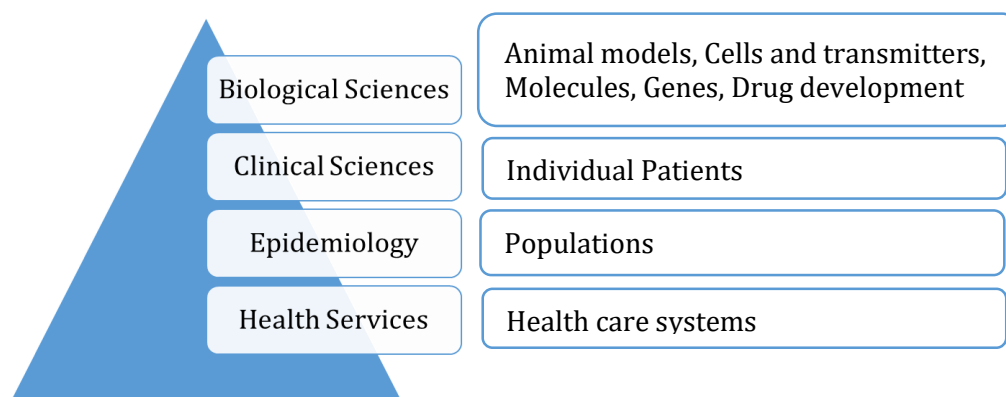
The science of making predictions about individual patients by counting clinical events in similar patients, using strong scientific methods for studies of groups of patients to ensure that the predictions are accurate. OR Clinical epidemiology is the application of epidemiological principles and methods to the practice of medicine

- Used as an aid to clinical decision making
- Lead to valid conclusions by avoiding being misled by systematic error (bias) and chance

Clinical Epidemiology is further could be defined as "Study/investigation /research/process of the distribution and determinants of health related states and event of individual and application of this findings to control the individual health problems." *Revised as definition of clinical epidemiology.*

Purpose

Is to develop and apply methods of clinical observation that will lead to valid conclusion by avoiding being misled by systematic error and the play of chance



Why learn clinical epidemiology (CE)

- Clinical training has been oriented toward the mechanisms of disease through study of biochemistry, anatomy, physiology and other traditional basic science
- Knowledge of the biology of disease should ordinarily be considered hypothesis, to be tested by clinical research -mechanism are only partly understood and many other factors in the genetic, physical and social environments affect outcome

- Clinical epidemiology – foundation on which modern medicine is practiced

Types of questions addressed by clinical epidemiology (Elements)

Clinical Questions

Issue	Question
Abnormality	Is the patient sick or well?
Diagnosis	How accurate are tests used to diagnose disease?
Frequency	How often does a disease occur?
Risk	What factors are associated with an increased risk of disease?
Prognosis	What are the consequences of having a disease?
Treatment	How does treatment change the course of disease?
Prevention	Does an intervention on well people keep disease from arising? Does early detection and treatment improve the course of disease?
Cause	What conditions lead to disease? What are the pathogenetic mechanisms of disease?
Cost	How much will care for an illness cost?

Principal of CE

- Every disease has definitive determinants
- Every disease has definitive etiological/causatives agents
- Clinical epidemiology is to foster methods of clinical observation and interpretation that lead to valid conclusion and better patient care
- Health outcome of particular concern to patients and those caring for them such as sign/symptoms, disability and death
- In most clinical situation the diagnosis. Prognosis and results of treatment are uncertain for an individual ie all diagnosis process and treatment based on prediction probability

Application of CE

- Is to develop and apply methods of clinical observation that will lead to valid conclusion by avoiding being misled by systematic error and the play of chance
- Identify the current situation of patient to choose the correct treatment with high probability
- Diagnosis, prognosis and treatment
- Prevention and Control the disease transmission including morbidity trend, mortality, complication/disability

Uses of CE

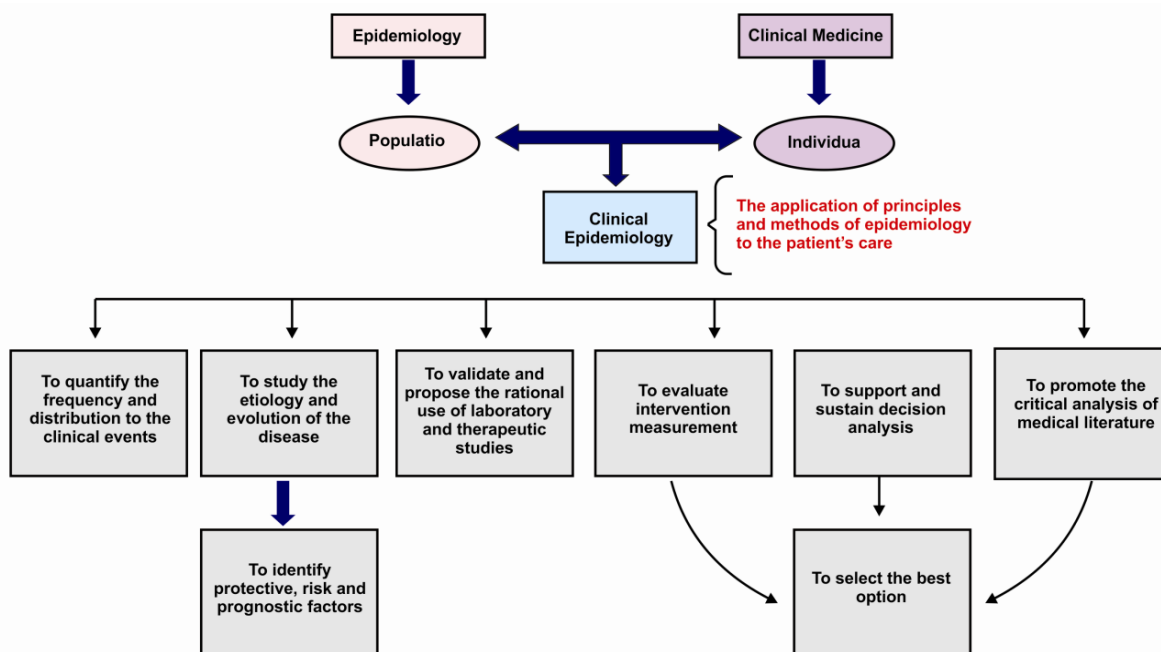
- Understanding of how bias and chance can affect the accuracy of observations in individual patients.
- Evaluation of the validity of test over diagnosis, prognosis, treatment and prevention.
- Knowing the strengths and weaknesses of randomized clinical trials, case-control studies, cohort studies (prospective and retrospective) and meta-analysis.
- Using practical strategies to judge the validity of clinical evidence synthesis (i.e. reviews).

- Comprehension of the meaning, uses and limits of the statistic power, the values of 'p' and the confidence interval, relative risk, attributable risk and NNH (number needed to harm).
- Knowing how to measure the patients' preferences.
- Comprehension and usage of the sensibility analysis and the cost-effectiveness analysis.

Public health context of clinical epidemiology

The current medical approach must contemplate the application of clinical epidemiology in the health establishments, patients; in general, epidemiologic research with its analytical designs and clinical trials allow the progress in treatments and managements

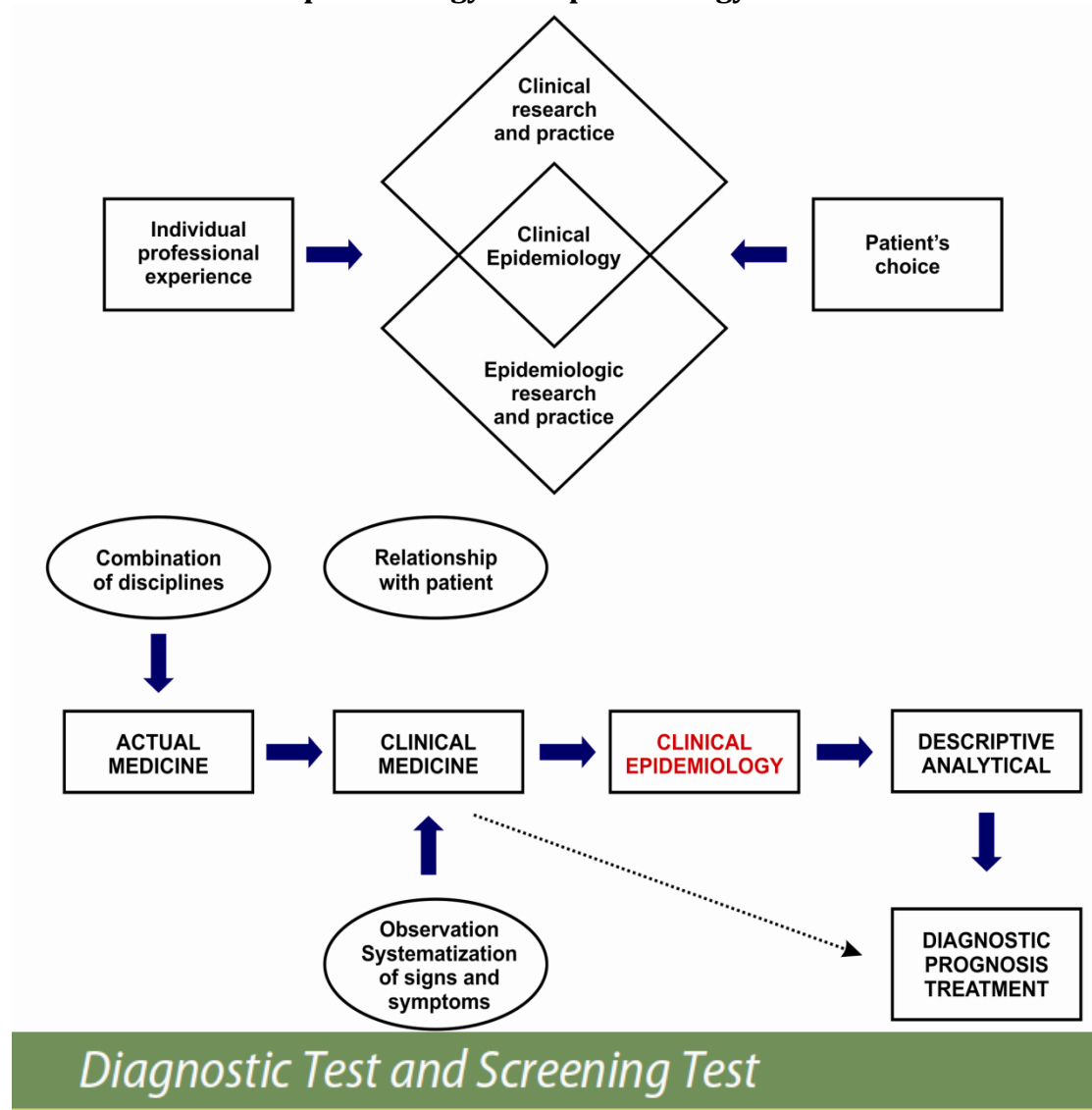
Clinical epidemiology has its application peak in the solution of treatment and management of diseases, contributing the identification of risk factors to certain illnesses and being to date the fundamental part of Evidence Based Medicine, and for that it becomes important the teaching of clinical epidemiology, for it will aid in the education of professionals with judicious capacity and rational use of the best alternatives in diagnostic and treatment.



Public Health keeps the cause-effect logic, and allows adding the determinants of health, such as lifestyles, culture and environment, to the scientific context that often is what characterizes the good or bad execution of an intervention.

Public Health is a new tendency that is joining efforts to offer the best information for an efficient politics decision making. When it is oriented to the research of a population's health issues whether in the community or the hospital sphere, it contributes remarkably to the solution of very different local and regional realities, thus making a progress in Public Health, especially in out developing countries

Relation of clinical epidemiology and epidemiology



Diagnostic Test and Screening Test

- A **diagnostic test** is used to determine the presence or absence of a disease when a subject shows signs or symptoms of the disease
- A **screening test** identifies asymptomatic individuals who may have the disease
- The diagnostic test is performed **after** a positive screening test to establish a definitive diagnosis

4.2. Screening

Definition

Screening people for disease or risk factors which predict disease – is motivated by the potential benefits of secondary prevention through early detection and treatment. 'Screening' is the process of detecting disease early, with the intention of intervening to halt its progression. This is linked to the idea of secondary prevention: interventions to stop further decline once a disease has begun but before symptoms appear. Secondary prevention requires a way to proactively detect disease as early as possible, perhaps even before the patient is aware of it. This is the role of screening.

This defined as "Screening is the process by which unrecognized disease or defects are identified, using tests which can be applied rapidly and on to large numbers of people. *Screening tests* distinguish apparently healthy people from those who probably have the disease." Screening is an initial examination; it is usually not diagnostic and requires appropriate investigative follow-up and treatment.

Aim of screening

Whilst we are busy and preoccupied with our daily lives, it is easy for us to neglect the importance and need for regular health screening. Many of the critical diseases that we commonly face today, such as heart disease and cancers, are often asymptomatic until the very late stages. Regular and comprehensive health screening enables us to identify your risk factors, detect diseases early and implement effective treatment strategies as soon as possible. We all know that "prevention is better than cure", and health screening is an important aspect of disease prevention and maintaining good health.

- To sort out those having the disease and those not having the disease from a group of apparently healthy individuals.

Uses of screening

1. Case detection (Prescriptive screening)

- It is also known as prescriptive screening.
- Defined as presumptive identification of unrecognized disease which does not arise from patients request. e.g. neonatal screening.
- Diseases sought include breast cancer, cervical cancer, diabetes, tuberculosis, HDN etc.
- Initiated by medical and public health personnel, they are under special obligation to make sure that appropriate treatment is started early.

2. Control of the disease (Prospective screening)

- Also known as prospective screening.
- People are examined for benefit of others. e.g. screening of immigrants from infectious diseases such as tuberculosis, syphilis, streptococcal infection etc.
- The screening programme may, by leading to early diagnosis permit more effective treatment and reduce the spread of infectious disease and/or mortality from the disease.

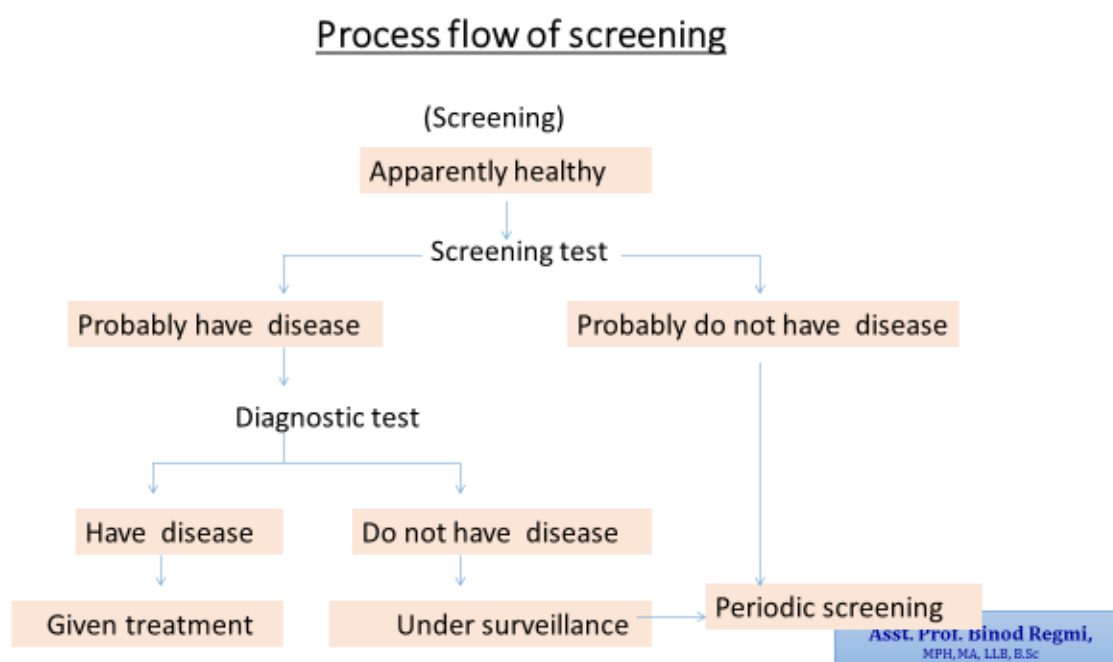
3. Research purpose

- Screening helps in obtaining basic knowledge about natural history of chronic diseases.
- Initial screening provides prevalence estimate and subsequent screening, an incidence figure.
- Where screening is done for research purposes, the investigator should inform the study participant that no follow up therapy will be available.

4. Educational purpose

- Screening programs provide opportunities for creating public awareness and for educating health professionals. eg screening for diabetes.

Process flow of screening



Diagnosis and Screening: Differences between screening and diagnostic tests

Screening tests are not diagnostic tests

The primary purpose of screening tests is to detect early disease or risk factors for disease in large numbers of apparently healthy individuals.

The purpose of a diagnostic test is to establish the presence (or absence) of disease as a basis for treatment decisions in symptomatic or screen positive individuals (confirmatory test). Some of the key differences are tabled below:

Differences between screening and diagnostic tests

Areas	Screening tests	Diagnostic tests
Purpose	To detect potential disease indicators	To establish presence/absence of disease

Areas	Screening tests	Diagnostic tests
Target population	Large numbers of asymptomatic, but potentially at risk individuals Ie. Done on apparently healthy people	Symptomatic individuals to establish diagnosis, or asymptomatic individuals with a positive screening test Ie. Done on sick people
	Done on groups	Done on individual cases
Test method	Simple, acceptable to patients and staff	maybe invasive, expensive but justifiable as necessary to establish diagnosis
Test result	Test results are final	Diagnosis is not final but based on other criteria such as history and clinical findings
Positive result threshold	generally chosen towards high sensitivity not to miss potential disease	Chosen towards high specificity (true negatives). More weight given to accuracy and precision than to patient acceptability
Purpose	to do community diagnosis, to launch a control programme	The purpose is to make a diagnosis in the patient to give treatment
Cost	Cheap, benefits should justify the costs since large numbers of people will need to be screened to identify a small number of potential cases	Higher costs associated with diagnostic test maybe justified to establish diagnosis.
Implemented by	Done by the epidemiologist	Done by the physician

Type of screening diagnostic test

1. Mass screening

Application of screening test to large, unselected population. Everyone in the group is screened regardless of the probability of having the disease or condition

- Means screening of the whole population or a sub group
- It is offered to all irrespective of particular risk individual may run of contracting disease in question.
- Indiscriminate mass screening is not a useful preventive measure unless backed up by suitable treatment that will reduce duration of illness or alter final outcome.

Example:

- visual defects in school children
- mammography in women aged 40 years or less
- newborn screening program in Japan

2. High –risk screening

- Screening is most productive if applied selectively to high risk groups, groups defined on epidemiological research.
- Diseases tend to aggregate in family so screening other members of family can detect additional cases.
- Epidemiologists have extended concept of screening to “risk factors.”

Example:

- screening fetus for Down’s syndrome in a mother who already has a baby with Down’s syndrome
- screening for familial cancers, HTN and DM
- screening for CA Cervix in low Socio economic status women
- screening for HIV in risk groups

3. Multipurpose screening

The screening of a population by more than one test done simultaneously to detect more than one disease

Example:

- Screening of pregnant women for VDRL, HIV, HBV by serological tests

4. Multiphase screening

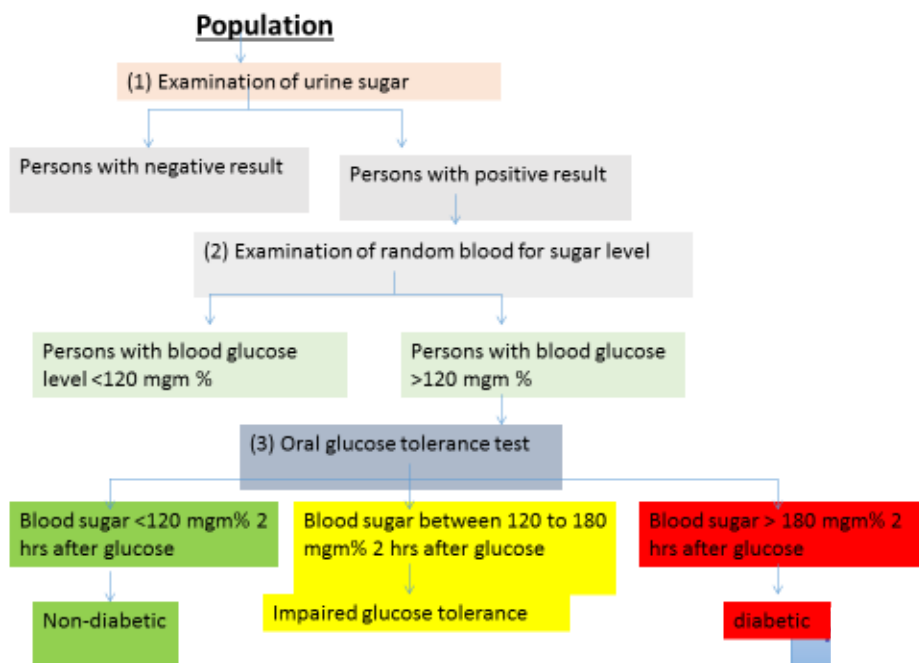
The screening in which various diagnostic procedures are employed during the same screening program.

- Defined as application of two or more screening test in combination to large number of people at one time than to carry out separate screening test for single diseases.
- Procedure includes health questionnaire, clinical examination and range of measurements and investigations. E.g. chemical and hematological tests on blood and urine, lung function assessment, audiometry etc.

Example:

- DM – FBS, Glucose tolerance test
- Sickle cell anemia – CBC, Hb electrophoresis

Diagrammatic examples of multiphase screening



5. Opportunistic screening

There is no accurate or precise diagnostic test for the disease and where the frequency of its occurrence in the population is small. The main objective is to detect disease and bring patients to treatment.

Example:

– RHD in children

Criteria of screening

- The condition being screened for is important in terms of public health – for example, it is serious and/or affects large numbers of people
- The natural history of the condition should be well understood
- There is a recognizable early stage of the condition
- There is an effective and acceptable treatment for the condition, so screening is likely to make a difference to its outcome
- There is a valid and reliable test for the condition that is acceptable to people being offered screening
- The screening programme is of good quality and cost-effective in the setting in which it is to be offered
- The information provided to people is unbiased; based on good evidence; and clear about possible harms (eg, over diagnosis leading to over-treatment) as well as potential benefits
- The chance of physical or psychological harm to those offered screening is likely to be less than the chance of benefit
- There are adequate facilities (health service provision) for the diagnosis and treatment of abnormalities detected by screening
- Good Prognosis

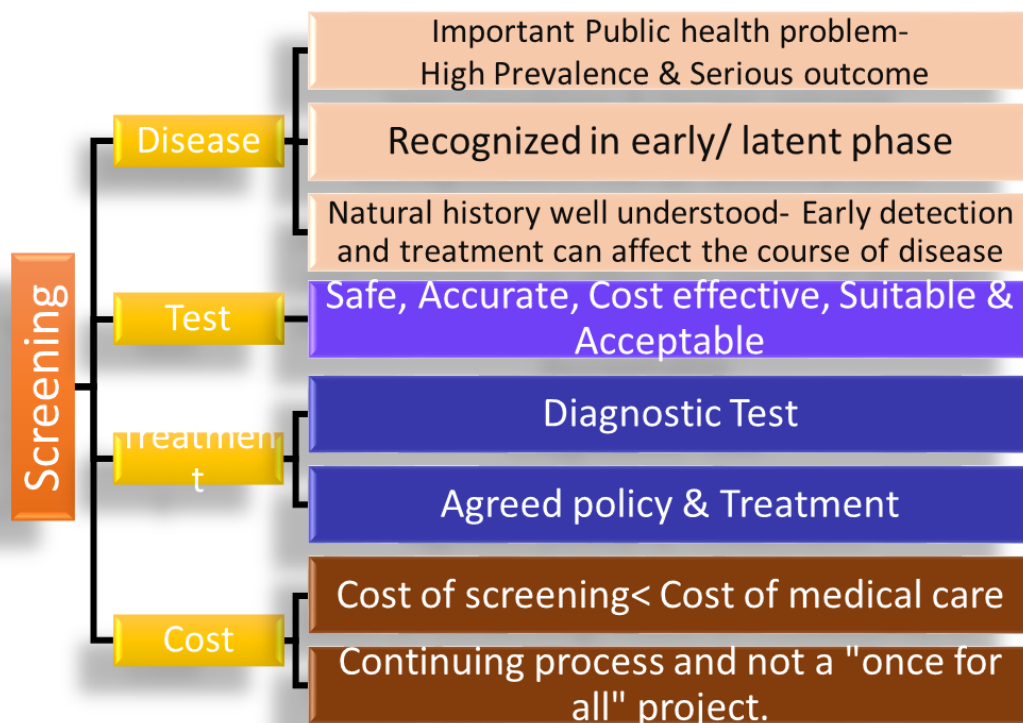
Furthermore the screening separating the criteria of screening on the following basis also

According to Disease

- Significant burden of disease
- Detectable and long preclinical stage of disease
- Adequately understood natural history of disease
- Appropriate test available for early detection of disease
- Facilities for diagnosis of disease
- Early detection of disease has outcome benefit
- Effective treatment available for disease
- Policy of screening program for disease

Screening test

- Inexpensive
- Acceptable
- Valid
- Reliable
- Yielding



Advantage and disadvantage of screening test

Advantages

- Screening makes it possible to detect many diseases at early stage

- Minimize the risk of developing disease
- Reduction of morbidity and mortality
- Low economic burden among the people who have chances of development of disease if not timely screened
- Treatment that begins prior to the appearance of symptoms makes it possible to avoid serious and permanent consequences
- For many diseases, early treatment can even enable completely normal development of people
- In general, examples, a single blood specimen and a single urine specimen will suffice to detect all the target diseases.

Disadvantages

- Occasionally, specimens may need to be taken a second time.
- Variant of the disease is rare or more difficult to diagnose, and naturally, this could be a source of concern for the parents.
- Even if treatment begins at an early age, complications of the disease could still arise.
- Despite its high efficacy, screening does have limits. For example, there exists a small chance that people does have a target disease that went undetected during the screening process.
- On the other hand, screening may detect other rare diseases not targeted by the test.
- It gives sometimes-false positive results that create unneeded worrying.
- Sometime, over diagnosis which result in unnecessary treatment.

Various screening programme in Nepal

School-based screening

Most public school systems in the United States screen students periodically for hearing and vision deficiencies and dental problems.

Cervical screening program

As part of the different organization conducted many screening project, Thousands of women around the country have been screened for cervical cancer and inflammatory diseases of the cervix during several screening camps organized in different districts.

Rheumatic fever screening Programme

With support of MOHP the National heart foundation, Gangalal Hospital organized the Rheumatic fever and effect in heard among the adolescent of different districts

DM and Hypertension screening program

Such type of screening programme was organized in government hospitals, community lavel and by different organization to identify the hypertension and DM

Low vision and refractive errors screening program

Refractive errors (myopia, hypermetropia, astigmatism, and presbyopia) result in an unfocussed image falling on the retina. Uncorrected refractive errors, which affect persons

of all ages and ethnic groups, are the main cause of visual impairment. Nepal Netra Jyoti Sangh periodically organized the screening program in Nepal.

Nepal Sickle Cell Disease Screening

The purpose of this project was to estimate the prevalence of SCD in the indigenous Tharu community. With a team of Nepali health professionals, the team drew blood from more than 2900 hundred individuals, performed blood smears, and analyzed the samples using microscopy. This was the largest health screening in Nepal's history. Individuals that screened positive for the trait were invited back to receive information and counselling on how to manage their symptoms and plan for their futures.

Community-Based Screening for Chronic Kidney Disease by BPKIHS

A community-based screening on, 3218 people ≥ 20 years were assessed by door-to-door survey in Dharan, Nepal. Health status, lifestyle habit, physical examination and blood pressure were evaluated. Spot urine was examined for proteins and glucose by dipstick. Fasting blood glucose and serum creatinine were measured in a subset of 1000 people and the prevalence of Chronic Kidney Disease was evaluated.

Etc....

Some Common Screening Tests

- Pap smear for cervical dysplasia or cervical cancer
- Fasting blood cholesterol for heart disease
- Fasting blood sugar for diabetes
- Blood pressure for hypertension
- Mammography for breast cancer
- PSA test for prostate cancer
- Faecal occult blood for colon cancer
- Ocular pressure for glaucoma
- PKU test for phenylketonuria in newborns
- TSH for hypothyroid and hyperthyroid

4.3. Diagnostic test function and confounding-Definition

A test that is applied to confirm or refute the existence of disease is called diagnostic test. In clinical medicine, studies concerned with misclassification of disease are usually called diagnostic test studies. The primary goal of such a study is to evaluate the performance of a test for diagnosing a disease condition of interest.

Suppose, for example, that the disease condition is deep vein thrombosis, or DVT. In a diagnostic study for DVT, the clinician targets only patients with a specific symptom, for example "acute leg swelling" and then performs both the diagnostic test, typically an

ultrasound, and the gold standard procedure, typically an x-ray venogram, on these patients. Here are the results of such a diagnostic test study in the form of a misclassification table.

Validity and Reliability,

For the statistical consultant working with social science researchers the estimation of reliability and validity is a task frequently encountered. Measurement issues differ in the social sciences in that they are related to the quantification of abstract, intangible and unobservable constructs. In many instances, then, the meaning of quantities is only inferred. Let us begin by a general description of the paradigm that we are dealing with. Most concepts in the behavioral sciences have meaning within the context of the theory that they are a part of. Each concept, thus, has an operational definition which is governed by the overarching theory. If a concept is involved in the testing of hypothesis to support the theory it has to be measured. So the first decision that the research is faced with is “how shall the concept be measured?” That is the *type of measure*. At a very broad level the type of measure can be observational, self-report, interview, etc.

These types ultimately take shape of a more specific form like observation of ongoing activity, observing video-taped events, self-report measures like questionnaires that can be open-ended or close-ended, Likert-type scales, interviews that are structured, semi-structured or unstructured and open-ended or close-ended. Needless to say, each type of measure has specific types of issues that need to be addressed to make the measurement meaningful, accurate, and efficient.

Another important feature is the *population* for which the measure is intended. This decision is not entirely dependent on the theoretical paradigm but more to the immediate research question at hand.

A third point that needs mentioning is the *purpose of the scale* or measure. What is it that the researcher wants to do with the measure? Is it developed for a specific study or is it developed with the anticipation of extensive use with similar populations?

Once some of these decisions are made and a measure is developed, which is a careful and tedious process, the relevant questions to raise are “how do we know that we are indeed measuring what we want to measure?” since the construct that we are measuring is abstract, and “can we be sure that if we repeated the measurement we will get the same result?”. The first question is related to validity and second to reliability. Validity and reliability are two important characteristics of behavioral measure and are referred to as psychometric properties.

It is important to bear in mind that validity and reliability are not an all or none issue but a matter of degree.

Validity:

Very simply, validity is the extent to which a test measures what it is supposed to measure. The question of validity is raised in the context of the three points made above, the form of the test, the purpose of the test and the population for whom it is intended. Therefore, we cannot ask the general question “Is this a valid test?”. The question to ask is “how valid is

this test for the decision that I need to make?” or “how valid is the interpretation I propose for the test?” We can divide the types of validity into logical and empirical.

- **Content Validity:**

When we want to find out if the entire content of the behavior/construct/area is represented in the test we compare the test task with the content of the behavior. This is a logical method, not an empirical one. Example, if we want to test knowledge on American Geography it is not fair to have most questions limited to the geography of New England.

- **Face Validity:**

Basically face validity refers to the degree to which a test appears to measure what it purports to measure.

- **Criterion-Oriented or Predictive Validity:**

When you are expecting a future performance based on the scores obtained currently by the measure, correlate the scores obtained with the performance. The later performance is called the criterion and the current score is the prediction. This is an empirical check on the value of the test – a criterion-oriented or predictive validation.

- **Concurrent Validity:**

Concurrent validity is the degree to which the scores on a test are related to the scores on another, already established, test administered at the same time, or to some other valid criterion available at the same time. Example, a new simple test is to be used in place of an old cumbersome one, which is considered useful, measurements are obtained on both at the same time. Logically, predictive and concurrent validation are the same, the term concurrent validation is used to indicate that no time elapsed between measures.

- **Construct Validity:**

Construct validity is the degree to which a test measures an intended hypothetical construct. Many times psychologists assess/measure abstract attributes or constructs. The process of validating the interpretations about that construct as indicated by the test score is construct validation. This can be done experimentally, e.g., if we want to validate a measure of anxiety. We have a hypothesis that anxiety increases when subjects are under the threat of an electric shock, then the threat of an electric shock should increase anxiety scores (note: not all construct validation is this dramatic!)

Reliability:

Research requires dependable measurement. Measurements are reliable to the extent that they are repeatable and that any random influence which tends to make measurements different from occasion to occasion or circumstance to circumstance is a source of measurement error. (Gay) Reliability is the degree to which a test consistently measures whatever it measures. Errors of measurement that affect reliability are random errors and errors of measurement that affect validity are systematic or constant errors.

Test-retest, equivalent forms and split-half reliability are all determined through correlation.

– **Test-retest Reliability:**

Test-retest reliability is the degree to which scores are consistent over time. It indicates score variation that occurs from testing session to testing session as a result of errors of measurement. Problems: Memory, Maturation, Learning.

– **Equivalent-Forms or Alternate-Forms Reliability:**

Two tests that are identical in every way except for the actual items included. Used when it is likely that test takers will recall responses made during the first session *and* when alternate forms are available. Correlate the two scores. The obtained coefficient is called the coefficient of stability or coefficient of equivalence. Problem: Difficulty of constructing two forms that are essentially equivalent.

Both of the above require two administrations.

– **Split-Half Reliability:**

Requires only one administration. Especially appropriate when the test is very long. The most commonly used method to split the test into two is using the odd-even strategy. Since longer tests tend to be more reliable, and since split-half reliability represents the reliability of a test only half as long as the actual test, a correction formula must be applied to the coefficient. Spearman-Brown prophecy formula.

Split-half reliability is a form of internal consistency reliability.

Rationale Equivalence Reliability:

Rationale equivalence reliability is not established through correlation but rather estimates internal consistency by determining how all items on a test relate to all other items and to the total test.

Internal Consistency Reliability:

Determining how all items on the test relate to all other items. Kuder-Richardson- α is an estimate of reliability that is essentially equivalent to the average of the split-half reliabilities computed for all possible halves.

Validity and Reliability of Diagnostic Test

VALIDITY IS THE ACCURACY OF A TEST.

RELIABILITY IS THE PRECISION OF A TEST.

ACCURACY: “how close is result of a test to its true value?”

PRECISION: “how close are the results of a test on repetition?”

Validity of diagnostic test

- An expression of the degree to which a test measures what it purports to measure.

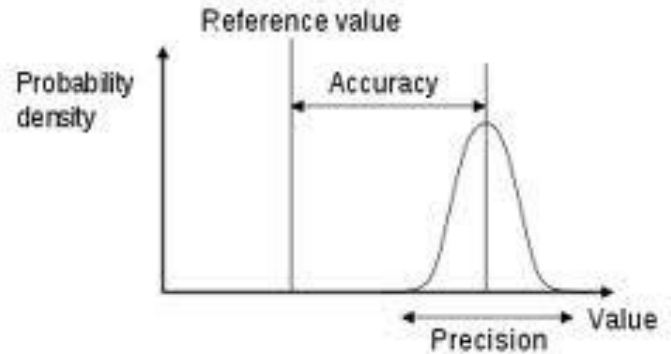
How to understand validity?

- ESTIMATE: a measurement or statement about the value of a quantity under study.
- An epidemiological estimate is the end product of the study design, the study conduct and the data analysis.

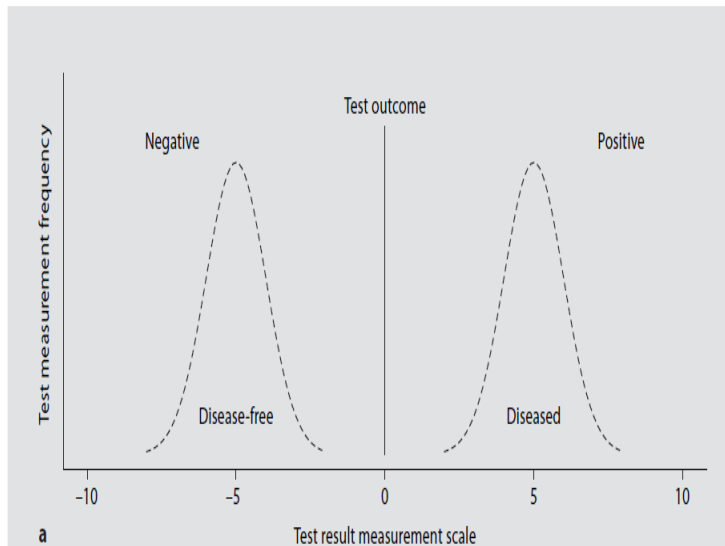
- Error is a false or mistaken result obtained in a study or experiment.
- Systematic error is one sided variation of measurements from the true value. It has a recognizable source.
- A study that has little systematic error is described as VALID.

COMPONENTS OF VALIDITY

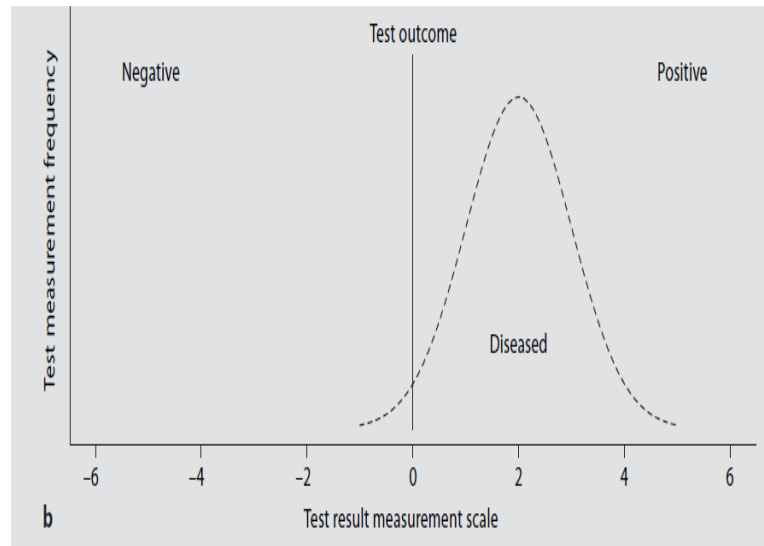
	Ds present	Ds absent
Test positive	Group (a) True Positive	Group (b) False Positive
Test negative	Group (c) False Negative	Group (d) True Negative



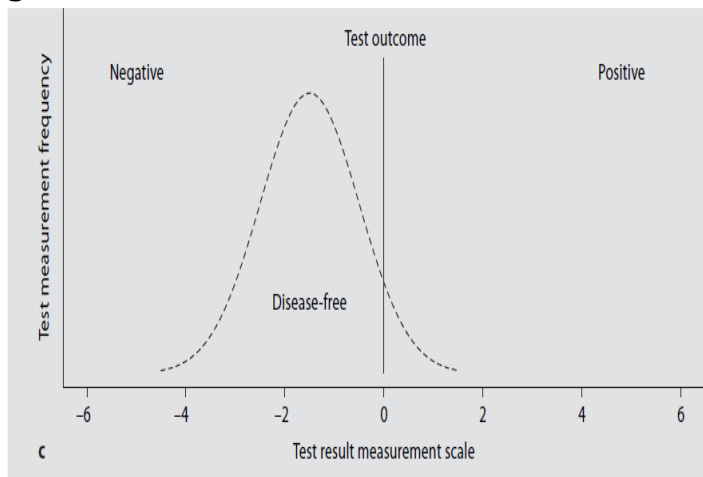
1



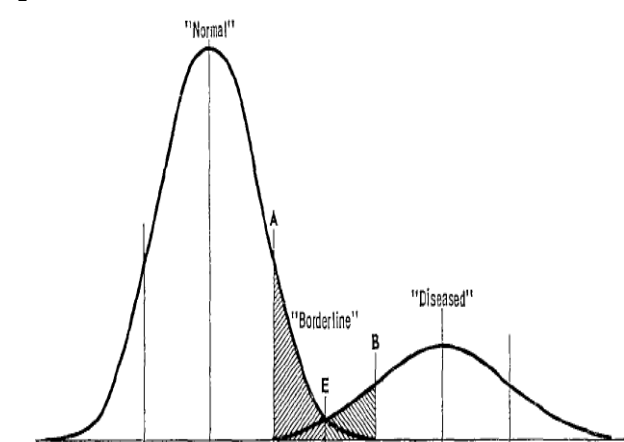
2



3



4



Validity and Reliability of diagnostic test is

1. Sensitivity
2. Specificity
3. Accuracy

Reliability of diagnostic test

Reproducibility, Repeatability, Reliability

- **Reproducibility, repeatability, reliability** all mean that the results of a test or measure are identical or closely similar each time it is conducted
- Because of variation in laboratory procedures, observers, or changing conditions of test subjects (such as time, location), a test may not consistently yield the same result when repeated
- Different types of variation
 - Intra-subject variation
 - Intra-observer variation
 - Inter-observer variation

Reproducibility is the ability of a test method to provide consistent results, as determined by estimates of precision, when applied to aliquots of the same samples tested in different laboratories, preferably located in distinct or different regions or countries using the identical assay (protocol, reagents and controls).

Repeatability is the level of agreement between results of replicates of a sample both within and between runs of the same test method in a given laboratory. Repeatability is estimated by evaluating variation in results of replicates.

Intra-Subject Variation

- **Intra-subject variation** is a variation in the results of a test conducted over (a short period of) time on the same individual
- The difference is due to the changes (such as physiological, environmental, etc.) occurring to that individual over that time period

Variation in Blood Pressure Readings: A 24-Hour Period

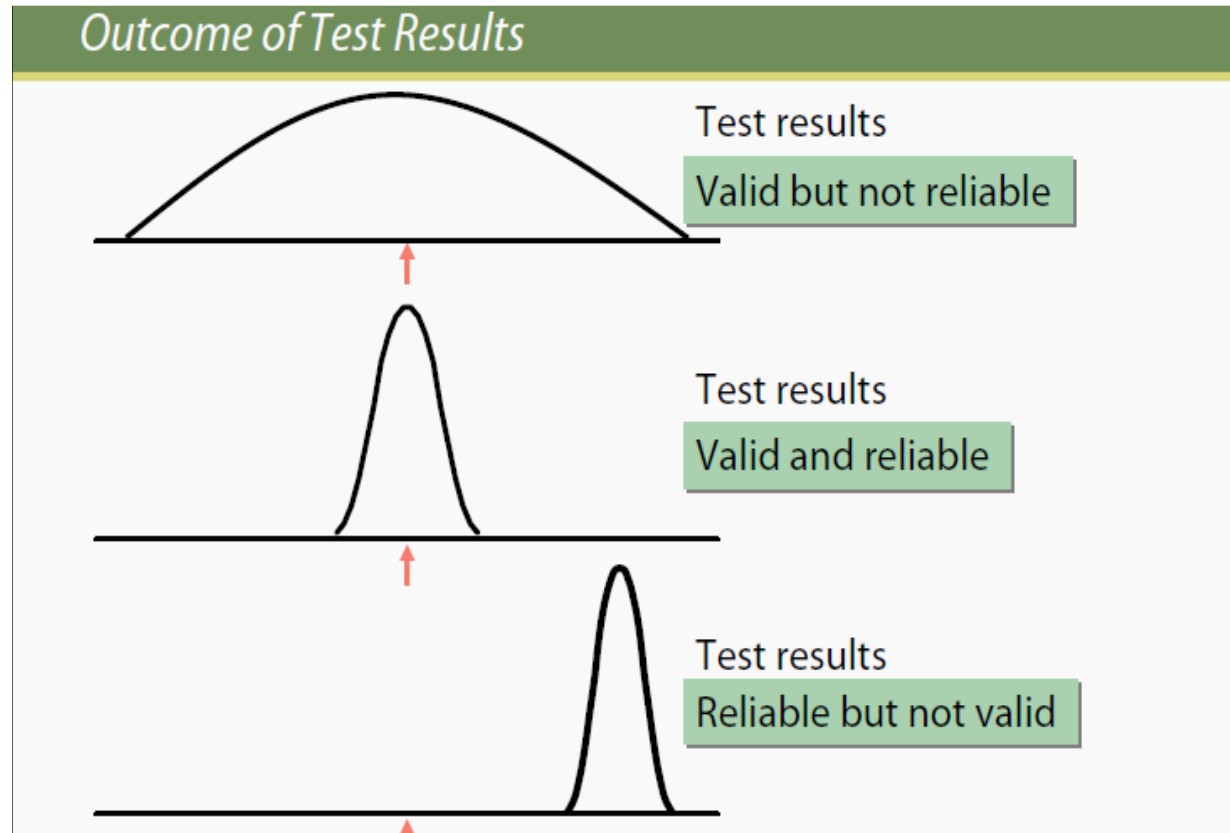
Blood Pressure (mm Hg)	Female 27 Years Old	Female 62 Years Old	Male 33 Years Old
Basal	110/70	132/82	152/109
Lowest hour	86/47	102/61	123/78
Highest hour	126/79	172/94	153/107
Casual	108/64	155/93	157/109

Inter-Observer and Intra-Observer Variation

- **Inter-observer variation** is a variation in the result of a test due to multiple observers examining the result (inter = between)
- **Intra-observer variation** is a variation in the result of a test due to the same observer examining the result at different times (intra = within)
- The difference is due to the extent to which observer(s) agree or disagree when interpreting the same test result

Examples of inter-observer and intra observer variation

- INTRA-OBSERVER VARIATION:
two readings of blood pressure measurement
- INTER-OBSERVER VARIATION:
chest x-ray films by different radiologists



4.4. Yield

Yield is the amount of unrecognized disease that is detected and brought to treatment as a result of screening.

$$\text{YIELD} = \text{TP} + \text{FP} / \text{TP} + \text{FP} + \text{TN} + \text{FN}$$

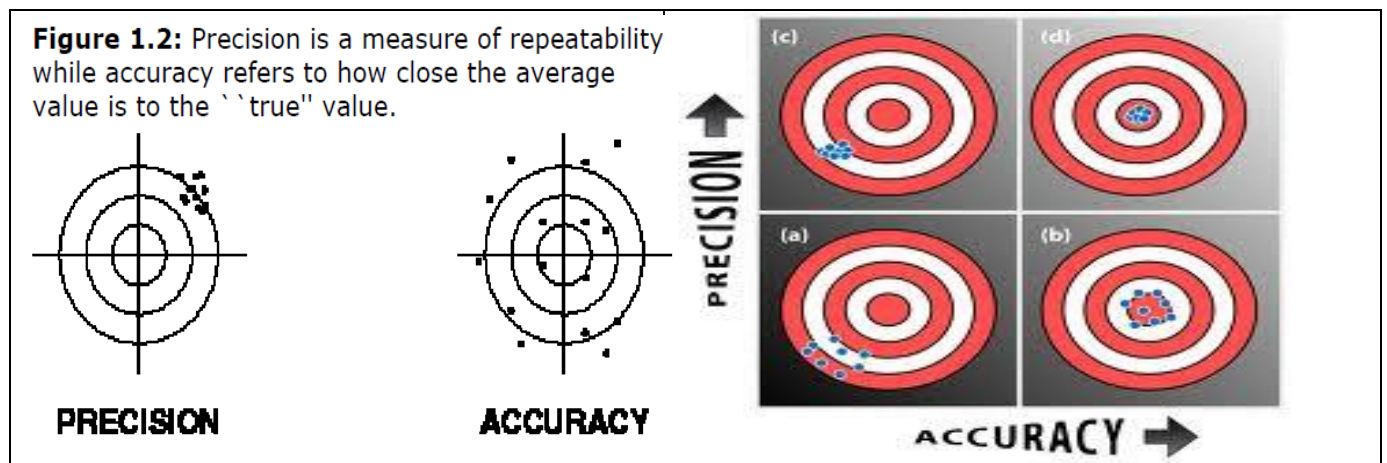
- It depends on prevalence of the disease and sensitivity of the screening test.
- Hence, yield of a screening test is high in high – risk screening

4.5. Precision

Accuracy is how close to "true" measurements are. Measuring devices or techniques can easily be inaccurate and lead to false measurements, and no matter how accurate a device may be, there is still a tolerance for error. No measurement is perfect. Accuracy must be accounted for in your results.

Precision is how consistent your results are for the same phenomena over several measurements, or how repeatable a device's (like a spring's) performance can be made. Precision as a measure of variation, must be accounted for in your calculations and results.

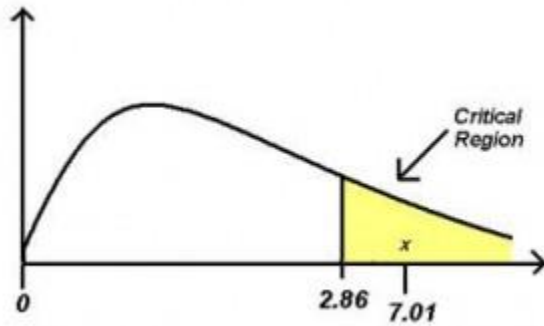
Accuracy is the extent to which a measurement reflects the true value. It is the tendency of test measurement to center around the true value. "Precise" means sharply defined or measured. Data can be very precise, but inaccurate. It would be precise but inaccurate to say that a meter equals 29.3748 inches. It would actually be more accurate to say that a meter equals a little over one yard, although that may not sound as impressive. An accurate measurement is one that is very close to the true value of the phenomenon we are observing. A precise measurement is one that has very little scatter: Repeat measurements give more or less the same value. If the measured data have high precision but poor accuracy, one may suspect that a systematic bias has been introduced, e.g., we are using an instrument where the zero position has not been set properly. If we do not know the expected value of a phenomenon but are trying to determine just that, it is obviously better to have accurate observations with poor precision than very precise, but inaccurate values, since the former will give a correct, but imprecise estimate while the latter will give a wrong, but very precise result.



4.6. Assumption of independence

What is the Assumption of Independence?

When multiple tests are used, as discussed above, the accuracy of the result depends on whether the additional information contributed by each test is somewhat independent of that already available from the preceding ones, i.e., the next test does not simply duplicate known information. In fact, this premise underlies the entire approach to predictive value we have discussed. However, it seems unlikely that the tests for most diseases are fully independent of one another. If the assumption that the tests are completely independent is wrong, cillculation of the probilibility of disease from several tests would tend to overestimate the tests' value.



The assumption of independence is a foundation for many statistical tests.

The assumption of independence is used for T Tests, in ANOVA tests, and in several other statistical tests. It's essential to getting results from your sample that reflect what you would find in a population. Even the smallest dependence in your data can turn into heavily biased results (which may be undetectable) if you violate this assumption.

A **dependence** is a connection between your data. For example, how much you have severity of infection of disease depends upon how many times you been exposed with cause.

Independence means there isn't a connection. For example, how much you severity of disease is not connected to what you have family members. The assumption of independence means that your data isn't connected in any way (at least, in ways that you haven't accounted for in your model).

There are actually two assumptions:

- The **observations between groups** should be independent, which basically means the groups are made up of different people. You don't want one person appearing twice in two different groups as it could skew your results.
- The observations **within each group** must be independent. If two or more data points in one group are connected in some way, this could also skew your data.

Based on above typical assumptions are:

- **Normality:** Data have a normal distribution (or at least is symmetric)
- **Homogeneity of variances:** Data from multiple groups have the same variance
- **Linearity:** Data have a linear relationship
- **Independence:** Data are independent

For example, let's say you were taking a snapshot of how many donuts people ate, and you took snapshots every morning at 9, 10, and 11 a.m. You might conclude that office workers eat 25% of their daily calories from donuts. However, you made the mistake of timing the snapshots too closely together in the morning when people were more likely to bring bags of donuts in to share (making them dependent). If you had taken your measurements at 7, noon and 4 p.m., this would probably have made your measurements independent.

What happens if violate the Assumption of Independence?

In simple terms, if you violate the assumption of independence, you run the risk that all of your results will be wrong. When these assumptions are violated the results of the analysis can be misleading or completely mistaken.

How do Avoid Violating the Assumption?

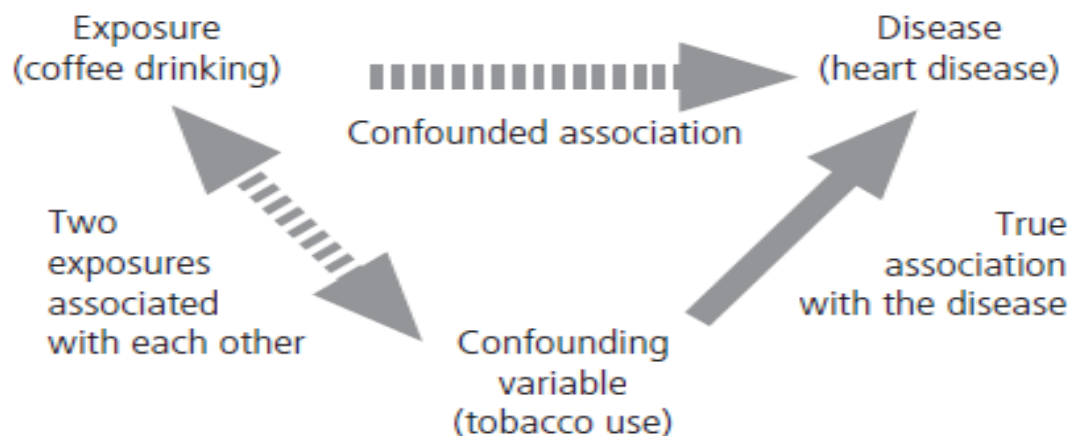
Unfortunately, looking at your data and trying to see if you have independence or not is usually difficult or impossible. The key to avoiding violating the assumption of independence is to make sure your data is independent while you are collecting it. If you aren't an expert in your field, this can be challenging. However, you may want to look at previous research in your area and see how the data was collected.

4.7. Confounding

Confounding is another major issue in epidemiological studies. In a study of the association between exposure to a cause (or risk factor) and the occurrence of disease, confounding can occur when another exposure exists in the study population and is associated both with the disease and the exposure being studied. A problem arises if this extraneous factor – itself a determinant or risk factor for the health outcome—is unequally distributed between the exposure subgroups. Confounding occurs when the effects of two exposures (risk factors) have not been separated and the analysis concludes that the effect is due to one variable rather than the other. To be a confounding factor, two conditions must be met as following.

In the example in Figure 3.10, confounding may be the explanation for the relationship demonstrated between coffee drinking and the risk of coronary heart disease, since it is known that coffee consumption is associated with tobacco use: people who drink coffee are more likely to smoke than people who do not drink coffee.

It is also well known that cigarette smoking is a cause of coronary heart disease. It is thus possible that the relationship between coffee drinking and coronary heart disease merely reflects the known causal association of tobacco use and heart disease. In this situation, smoking confounds the apparent relationship between coffee consumption and coronary heart disease because smoking is correlated with coffee drinking and is a risk factor even for those who do not drink coffee.



The control of confounding

Several methods are available to control confounding, either through study design or during the analysis of results. The methods commonly used to control confounding in the design of an epidemiological study are:

- Randomization
- Restriction
- Matching.

At the analysis stage, confounding can be controlled by:

- Stratification
- Statistical modeling.

Randomization

In experimental studies, randomization is the ideal method for ensuring that potential confounding variables are equally distributed among the groups being compared. The sample sizes have to be sufficiently large to avoid random maldistribution of such variables. Randomization avoids the association between potentially confounding variables and the exposure that is being considered.

Restriction

One way to control confounding is to limit the study to people who have particular characteristics. For example, in a study on the effects of coffee on coronary heart disease, participation in the study could be restricted to nonsmokers, thus removing any potential effect of confounding by cigarette smoking.

Matching

Matching is used to control confounding by selecting study participants so as to ensure that potential confounding variables are evenly distributed in the two groups being compared. For example, in a case-control study of exercise and coronary heart disease, each patient with heart disease can be matched with a control of the same age group and sex to ensure that confounding by age and sex does not occur. Matching has been used extensively in case-control studies but it can lead to problems in the selection of controls if the matching criteria are too strict or too numerous; this is called overmatching.

Matching can be expensive and time-consuming, but is particularly useful if the danger exists of there being no overlap between cases and controls, such as in a situation where the cases are likely to be older than the controls.

Stratification and statistical modelling

In large studies it is usually preferable to control for confounding in the analytical phase rather than in the design phase. Confounding can then be controlled by stratification, which involves the measurement of the strength of associations in well defined and homogeneous categories (strata) of the confounding variable. If age is a confounder, the association may be measured in, say, 10-year age groups; if sex or ethnicity is a confounder, the association is measured separately in men and women or in the different ethnic groups. Methods are

available for summarizing the overall association by producing a weighted average of the estimates calculated in each separate stratum.

Although stratification is conceptually simple and relatively easy to carry out, it is often limited by the size of the study and it can not help to control many factors simultaneously, as is often necessary. In this situation, multivariate statistical modeling is required to estimate the strength of the associations while controlling for several confounding variables simultaneously; a range of statistical techniques is available for these analyses.

4.8. Normality and Abnormality

Introduction

The first priority in any clinical consultation is to determine whether the patient's symptoms, signs or diagnostic test results are normal or abnormal. This is necessary before any further investigations or treatment. It would be easy if a clear distinction could always be made between measurements of normal and abnormal people.

There are three ways of distinguishing results in such a distribution:

Normality

- a. normal as common

Abnormality

- b. abnormal as associated with disease
- c. abnormal as treatable.

4.8.1 Normality

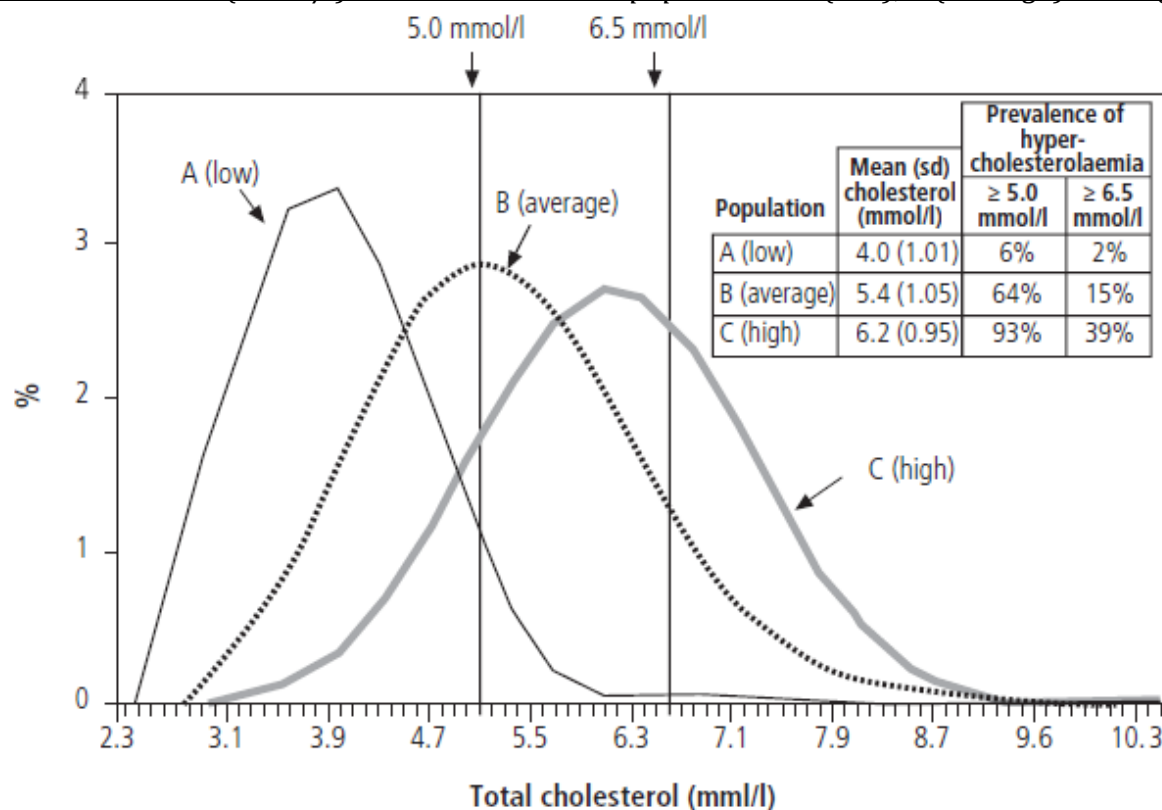
This definition classifies values that occur frequently as normal and those that occur infrequently as abnormal. We assume that an arbitrary cut-off point on the frequency distribution (often two standard deviations above or below the mean) is the limit of normality and consider all values beyond this point abnormal. This is called an operational definition of abnormality. If the distribution is in fact Gaussian (normal in the statistical sense), we would identify 2.5% of the population as abnormal by using this cut-off. An alternative approach, which does not assume a statistically normal distribution, is to use percentiles: we can consider that the 95th percentile point is the dividing line between normal and abnormally high values, thus classifying 5% of the population as abnormal. (R bonita)

4.8.2 Abnormality

- a. **abnormal as associated with disease**

The distinction between normal and abnormal can be based on the distribution of the measurements for both healthy and diseased people, and we can attempt to define a cut-off point that clearly separates the two groups. A comparison of two frequency distributions often shows considerable overlap –as illustrated by serum cholesterol distributions for people with and without coronary heart disease. Choosing a cut-off point that neatly separates cases from non-cases is clearly impossible as below fig.

Total cholesterol (mmol/l) distribution in three population : A (low), B (average) and C (high)



There are always some healthy people on the abnormal side of the cut-off point, and some true cases on the normal side. These two types of classification error can be expressed quantitatively in terms of the sensitivity and specificity of a test.

- Sensitivity is the proportion of truly diseased people who are categorized as abnormal by the test.
- Specificity is the proportion of truly normal people categorized as normal by the test. A balance always has to be struck between sensitivity and specificity; increasing one reduces the other.

b. Abnormal as treatable.

These difficulties in distinguishing accurately between normal and abnormal have led to the use of criteria determined by evidence from randomized controlled trials, which can be designed to detect the point at which treatment does more good than harm. Unfortunately, many treatment decisions have to be made in the absence of such evidence.

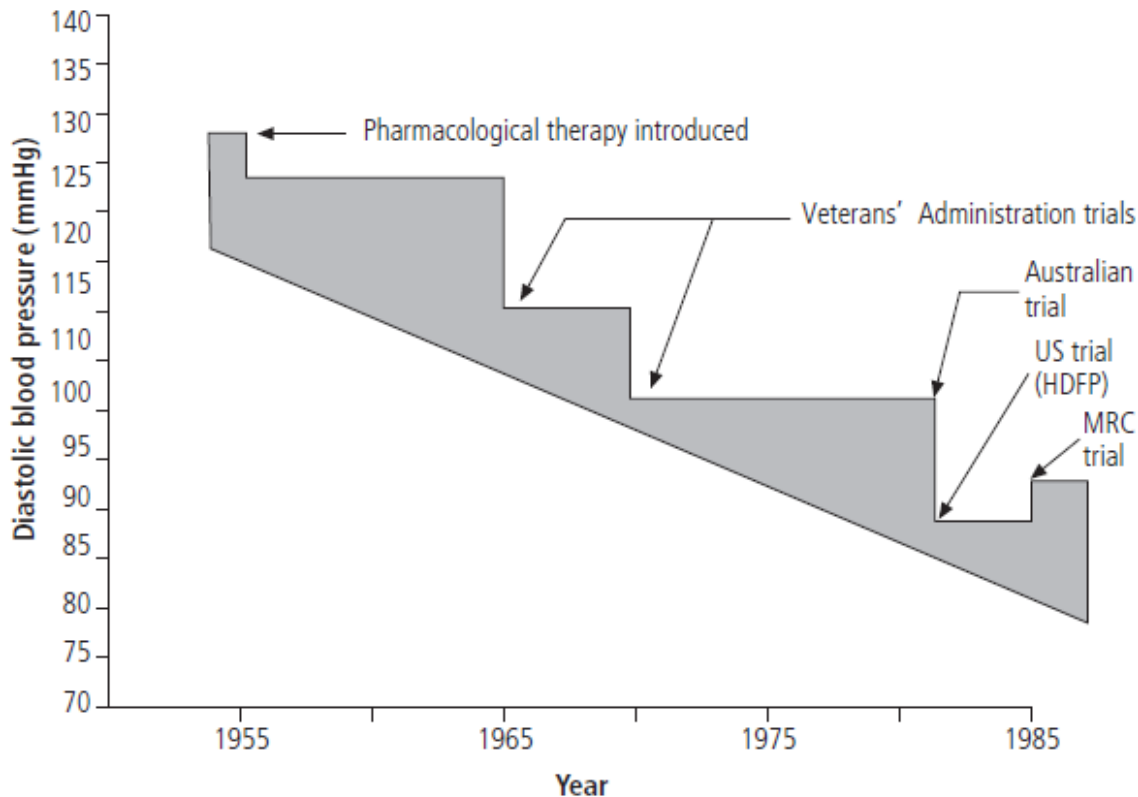
The treatment of high blood pressure is a good example – early clinical trials provided firm evidence that treating very high diastolic blood pressure (≥ 120 mmHg) was beneficial. Subsequent trials have indicated that the benefits of treatment outweigh the problems at lower levels of diastolic pressure, perhaps as low as 90 mmHg.

However, such trials are usually not designed to account for other risk factors or the cost of treatment. More sophisticated cost-effectiveness studies may make it possible to factor the economic consequences of treatment into clinical decisions.

We could then determine blood pressure levels at which treatment makes economic as well as medical sense for men and women in specific groups at risk. Treating a young woman with a diastolic blood pressure of 90 mmHg, who is at low overall risk of cardiovascular disease, will be much less cost-effective than treating an older man with a diastolic blood pressure of 90 mmHg, who has a much greater risk of cardiovascular disease. However, if the treatment of the young woman has no negative side-effects for her except the cost, she may choose to pay for the treatment herself.

What is considered to be worth treating changes with time; Figure 8.2 shows the changing definition of treatable levels of blood pressure. As clinical trials gather new evidence, treatment recommendations tend to change.

Figure 8.2. Treatment of hypertension: changing criteria over time



However, each time that a new cut-off point is proposed, we need to consider the logistic and cost implications. For example, if we take an evidence-based approach to treating people with mildly elevated blood pressure, we should be more concerned with assessing patients' absolute (or baseline) risk of cardiovascular disease, and place less emphasis on their actual blood pressure.³ Such risk prediction can help clinicians communicate with patients As box 8.1

Box 8.1. Risk prediction

Risk prediction (defining absolute risk of an event over a specified time period) provides clinicians with absolute measures of the effects of treatment and assists them in helping individuals with treatment decisions. Risk prediction charts can be used to account for multiple risk factors⁴. For example, a 5-year cardiovascular disease risk – of fatal or non-fatal events – is determined largely by a person's sex, age, diabetic status, smoking history, systolic blood pressure and total cholesterol. An individual's overall cardiovascular risk can be computed from a risk prediction chart. See http://www.nzgg.org.nz/guidelines/CVD_Risk_Chart.pdf for an example.

4.9. ROC Curve**Receiver Operating Characteristic Curves**

For a given diagnostic test, the true positive rate (TPR) against false positive rate (FPR) can be measured, where

$$\text{TPR} = \text{TP} / (\text{TP} + \text{FN})$$

And

$$\text{FPR} = \text{FP} / (\text{FP} + \text{TN})$$

As we can see from the above equations, TPR is equivalent to sensitivity and FPR is equivalent to (1 – specificity). All possible combinations of TPR and FPR compose a ROC space. One TPR and one FPR together determine a single point in the ROC space, and the position of a point in the ROC space shows the trade off between sensitivity and specificity, i.e. the increase in sensitivity is accompanied by a decrease in specificity. Thus the location of the point in the ROC space depicts whether the diagnostic classification is good or not. In an ideal situation, a point determined by both TPR and FPF yields a coordinates (0, 1), or we can say that this point falls on the upper left corner of the ROC space. This idea point indicates the diagnostic test has a sensitivity of 100% and specificity of 100%. It is also called perfect classification. Diagnostic test with 50% sensitivity and 50% specificity can be visualized on the diagonal determined by coordinate (0, 0) and coordinates (0, 1). Theoretically, a random guess would give a point along this diagonal. A point predicted by a diagnostic test fall into the area above the diagonal represents a good diagnostic classification, otherwise a bad prediction. A graphic presentation of what described above is shown in figure 1.

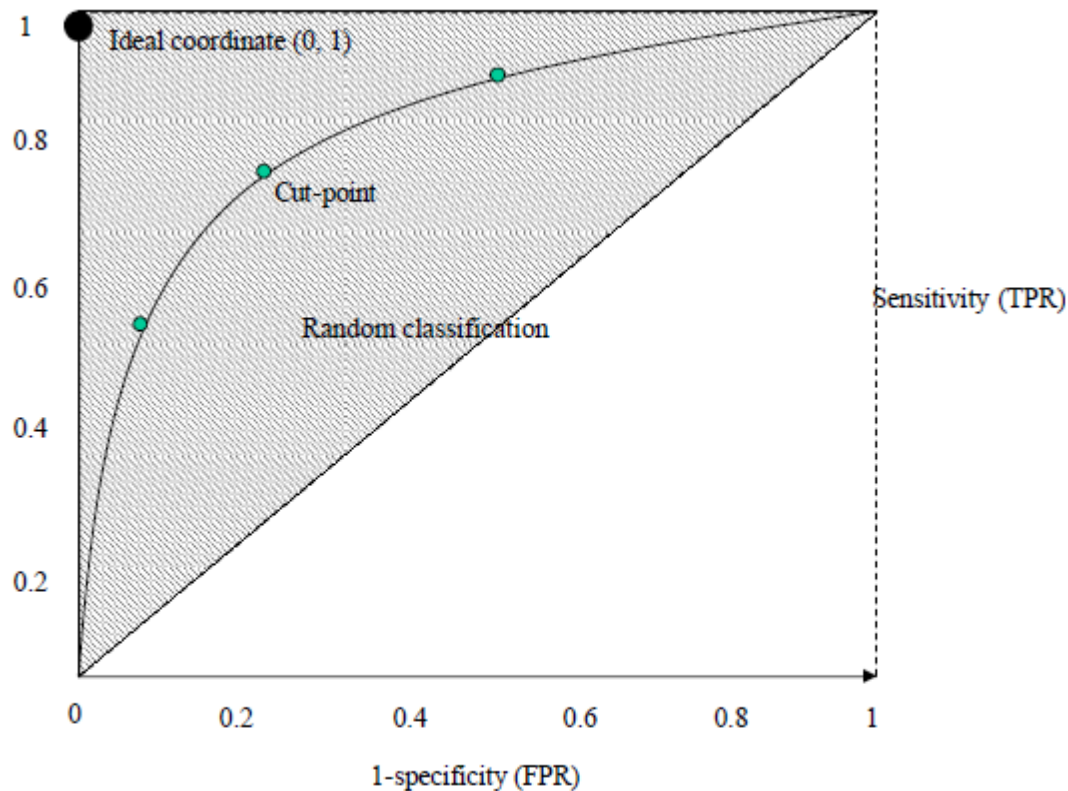


Figure 1: ROC Space: shadow area represents better diagnostic classification

ROC space, ROC curve is often plotted by using true positive rate (TPR) against false positive rate (FPR) for different cut-points of a diagnostic test, starting from coordinate (0, 0) and ending at coordinate (1, 1). FPR ($1 - \text{specificity}$) is represented by x-axis and TPR (sensitivity) is represented by y-axis. Thus, ROC curve is a plot of a test's sensitivity vs. ($1 - \text{specificity}$) as well. The interpretation of ROC curve is similar to a single point in the ROC space, the closer the point on the ROC curve to the ideal coordinate, the more accurate the test is. The closer the points on the ROC curve to the diagonal, the less accurate the test is.

In addition,

- (1) the faster the curve approach the ideal point, the more useful the test results are;
- (2) the slope of the tangent line to a cut-point tells us the ratio of the probability of identifying true positive over true negative, i.e. likelihood ratio (LR) for the test value: $LR = \text{sensitivity} / (1 - \text{specificity})$, if the ratio is equal to 1, the selected cut-point doesn't add additional information to identify true positive result. If the ratio is greater than 1, the selected cut-point help identify true positive result. If the ratio is less than 1, it decreases disease likelihood
- (3) the area under ROC curve (AUC) provides a way to measure the accuracy of a diagnostic test. The larger area, the more accurate the diagnostic test is. AUC of ROC curve can be measured by the following equation, Where $t = (1 - \text{specificity})$ and ROC (t) is sensitivity.

A single cut-point of a diagnostic test defines one single point in the ROC space; however, different possible cut points of a diagnostic test determine a curve in ROC space, which is also called ROC curve. Like a single point in the If you plot the sensitivity or “hits” versus (1-specificity) or “false alarms” that result from selecting different cut offs for the diagnostic test results, you generate a useful picture of the test’s accuracy that is called an “ROC curve1.” ROC curves nicely display the trade-offs of using one or more cut offs for the test. We show one for our BNP in below figure.

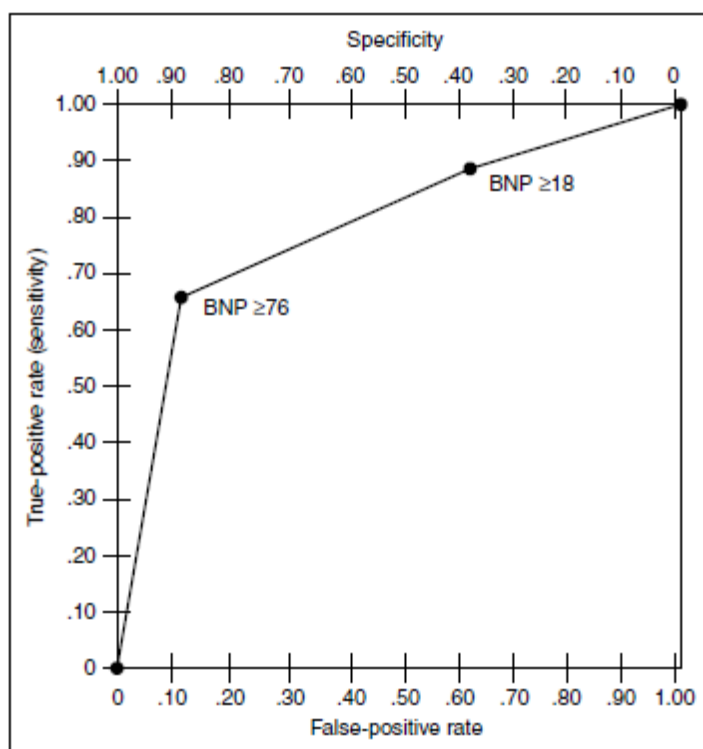


FIGURE 8-2 A receiver operating characteristic curve for B-type natriuretic peptide (BNP) as a diagnostic test for left ventricular dysfunction.

An ROC curve has some useful properties:

- It illustrates the performance of a dichotomous diagnostic test when you select different cut-points to distinguish “normal” from “abnormal” results.
- It demonstrates the fact that any increase in sensitivity will be accompanied by a decrease in specificity, and vice versa.
- The closer the curve gets to the upper left corner of the display, the more the overall accuracy of the test.
- The closer the curve comes to the 45-degree diagonal of the ROC space, the less accurate the test. At 45 degrees, the test adds no diagnostic information at all.
- The slope of the tangent at a cut-point gives the LR for that value of the test.
- The area under the curve provides an overall measure of a test’s accuracy. This property comes in handy when you are trying to decide which of two competing tests for the same target disorder is the better one. We’ll come back to that when we discuss the analysis of a diagnostic test study.

4.10. Description of performance of diagnostic and calculation

4.10.1. Gold Standard

Assessment of the test's accuracy rests on its relationship to some way of knowing whether the disease is truly present or not—a sounder indication of the truth often referred to as the "gold standard." As it turns out, the gold standard is often elusive. Sometimes the standard of accuracy is itself a relatively simple and inexpensive test, such as a throat culture for group A (Beta-hemolytic streptococcus) to validate the clinical impression of strep throat or an antibody test for human immunodeficiency virus. However, this is usually not the case. More often, one must turn to relatively elaborate, expensive, or risky tests to be certain whether the disease is present or absent. Among these are biopsy, surgical exploration, and of course, autopsy. In the gold standard means universally acceptable test for any kind of disease screening which have high level of sensitivity and specificity.

For diseases that are not self-limited and ordinarily become overt in a matter of a few years after they are first suspected, the results of follow-up can serve as a gold standard. Most cancers and chronic, degenerative diseases fall into this category. For them, validation is possible even if on the spot confirmation of a test's performance is not feasible because the immediately available gold standard is too risky, involved, or expensive. Some care must be taken in deciding the length of the follow-up period, which must be long enough for the disease to manifest but not so long that cases can arise after the original testing.

Because it is almost always more costly and more dangerous to use these more accurate ways of establishing the truth, clinicians and patients prefer simpler tests to the rigorous gold standard, at least initially. Chest x-rays and sputum smears are used to determine the nature of pneumonia, rather than lung biopsy with examination of the diseased lung tissue. Similarly, electrocardiograms and serum enzymes are often used to establish the diagnosis of acute myocardial infarction, rather than catheterization or imaging procedures. The simpler tests are used as proxies for more elaborate but, more accurate ways of establishing the presence of disease, with the understanding that some risk of misclassification results. This risk is justified by the safety and convenience of the simpler tests. But simpler tests are only useful when the risks of misclassification are known and found to be acceptably low. This requires sound data that compare their accuracy to an appropriate standard.

Gold Standard: an external source of truth regarding the disease status of each individual in the population. Gold standard is a benchmark and its results are considered definitive. Examples

- Infections – CULTURE
- Cancers – BIOPSY
- Drug testing – RANDOMIZED CONTROLLED TRIAL
- Cause of death – AUTOPSY

Definition, notation and calculation of test performance characteristics

4.10.2. Sensitivity

Ability of test to give positive results in a group of persons with the disease or probability of a positive test in people with the disease. A highly sensitive test will result in the correct identification of most individuals who have the disease. It will result in mislabeling as negative only a few who have the disease.

Number who test positive with disease (a)

Number with disease ($a + c$)

Sensitivity is the probability of testing positive given the presence of disease.

4.10.3. Specificity

Ability of the test to give a negative results in a group of persons without disease or probability of a negative test in people without the disease. A highly specific test will result in correct identification of most individuals who are free of disease.

Number who test negative without disease (d)

Number without disease ($b + d$)

Specificity is the probability of testing negative given the absence of disease.

For example,

The sensitivity of mammography for detecting breast cancer among women over 50 years old is about 85%. The interpretation of this sensitivity value is, “among women with biopsy proven breast cancer, the chance of having a positive mammogram is 85%.”

The specificity of mammography for detecting breast cancer among women over 50 years old is about 95%. The interpretation of this specificity value is, “among women with biopsy proven absence of breast cancer, the chance of having a negative mammogram is 95%.”

Sensitivity and specificity characteristics generally remain consistent across different populations, or may vary to only a small degree. However, sensitivity and specificity characteristics do not provide important clinical information for individual patients. In the mammography example, women typically would not be interested in their probability of having a positive mammogram after they are diagnosed with breast cancer. Instead, they would like to know the opposite information, specifically, what is their chance of having breast cancer given a positive or negative mammography result? To answer this more clinically relevant question, two additional characteristics of screening tests are needed.

Uses of sensitive test

- A sensitive test should be chosen when there is an important penalty for missing a disease. The diseases which are dangerous but treatable e.g. tuberculosis, syphilis etc.
- Screening in order to discover the disease e.g. during periodic health examination.

Uses of specific test

- It is useful to confirm the diagnosis that has been suggested by other data.
- Highly specific tests are particularly needed when false positive results can harm patient physically, emotionally, or financially.

		The Truth		
		Has the disease	Does not have the disease	
Test Score:	Positive	True Positives (TP) a	False Positives (FP) b	$PPV = \frac{TP}{TP + FP}$
	Negative	False Negatives (FN) c	True Negatives (TN) d	

		Sensitivity	Specificity
		$\frac{TP}{TP + FN}$	$\frac{TN}{TN + FP}$
Or,		$\frac{a}{a + c}$	$\frac{d}{d + b}$

4.10.4. Positive predictive value

Proportion of disease individuals among those with positive test results or probability of the person having the disease when the test is positive

Number who test positive with disease (a)

Number who test positive (a + b)

Positive predictive value is the probability of disease given a positive test result.

4.10.5. Negative predictive value

Proportion of individuals without disease among those negative test results or probability of the person not having the disease when the test is negative

Number who test negative without disease (d)

Number who test negative (c + d)

Negative predictive value is the probability of no disease given a negative test result.

For example

The positive predictive value of mammography is 10% in a low-risk patient population. The interpretation of this positive predictive value is, “among low-risk women who have a positive mammogram, the probability of breast cancer is 10%.”

The negative predictive value of mammography in this same population is 98%. The interpretation of this negative predictive value is, “among low-risk women with a negative mammogram, the probability of having breast cancer is 2%.”

(Unlike specificity and sensitivity, positive and negative predictive values depend on the prevalence of disease in the screened population. The relationships between sensitivity, specificity, positive and negative predictive values are illustrated in the following examples.)

False negative rate

False positive rate

If not consideration on false negative and false positives, the following problem may arise among the persons

FN: false reassurance

- ignoring of disease signs and symptoms
- postponement of treatment
- detrimental to overall health

FP: further testing

- discomfort, inconvenience, anxiety
- burden on health facilities
- emotional trauma
- difficulty in “de-labeling”

4.10.6. Likelihood Ratio (LR)

The Likelihood Ratio (LR) is the likelihood that a given test result would be expected in a patient with the target disorder compared to the likelihood that that same result would be expected in a patient without the target disorder.

For example, you have a patient with anaemia and a serum ferritin of 60mmol/liter and you find in an article that 90 per cent of patients with iron deficiency anaemia have serum ferritins in the same range as your patient (= sensitivity) and that 15 per cent of patients with other causes for anaemia have serum ferritins in the same range as your patient (1 – specificity).

This means that your patient's result would be six times as likely (90/15) to be seen in someone with, as opposed to someone without, iron deficiency anaemia, and this is called the LR for a positive test result.

The LR is the probability of a given test result in a patient with the target disorder divided by the probability of that same result in a person without the target disorder. The components of the LR are calculated vertically, and like the sensitivity and specificity are immune to the prevalence.

The LR of a positive test result:

LR+ = probability that an individual with the target disorder has a positive test probability than an individual without the target disorder has a positive test

In other words, LR+ = True positivity rate / False positivity rate

Which is the same as sensitivity / 1- specificity

The LR of a negative test result:

LR- = probability that an individual with the target disorder has a negative test probability than an individual without the target disorder has a negative test

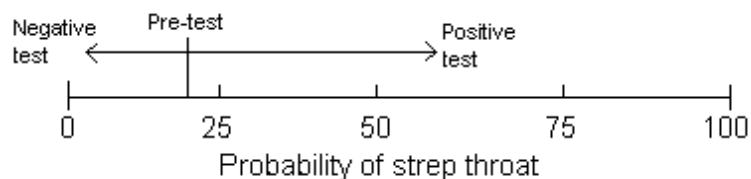
In other words, LR- = True negative rate / False negative rate

Which is the same as Specificity / 1- sensitivity

$$LR+ = \text{sensitivity} / (1-\text{specificity}) = (a/(a+c)) / (b/(b+d))$$

$$LR- = (1-\text{sensitivity}) / \text{specificity} = (c/(a+c)) / (d/(b+d))$$

When we decide to order a diagnostic test, we want to know which test (or tests) will best help us rule-in or rule-out disease in our patient. In the language of clinical epidemiology, we take our initial assessment of the likelihood of disease ("pre-test probability"), do a test to help us shift our suspicion one way or the other, and then determine a final assessment of the likelihood of disease ("post-test probability"). Take a look at the diagram below, which graphically illustrates this process of "revising the probability of disease".



Likelihood ratios tell us how much we should shift our suspicion for a particular test result. Because tests can be positive or negative, there are at least two likelihood ratios for each test. The "positive likelihood ratio" (LR+) tells us how much to increase the probability of disease if the test is positive, while the "negative likelihood ratio" (LR-) tells us how much to decrease it if the test is negative. The formula for calculating the likelihood ratio is:

$$\text{LR} = \frac{\text{Probability of an individual **with** the condition having the test result}}{\text{Probability of an individual **without** the condition having the test result}}$$

Thus, the positive likelihood ratio is:

$$\text{LR+} = \frac{\text{Probability of an individual **with** the condition having a positive test}}{\text{Probability of an individual **without** the condition having a positive test}}$$

Similarly, the negative likelihood ratio is:

$$\text{LR-} = \frac{\text{Probability of an individual **with** the condition having a negative test}}{\text{Probability of an individual **without** the condition having a negative test}}$$

You can also define the LR+ and LR- in terms of sensitivity and specificity:

$$\text{LR+} = \frac{\text{sensitivity}}{1 - \text{specificity}}$$

$$\text{LR-} = \frac{1 - \text{sensitivity}}{\text{specificity}}$$

(Of course, if you're using sensitivity and specificity on a scale of 0 to 100 instead of 0 to 1, the equations would be sensitivity / (100-specificity) and (100-sensitivity)/specificity, respectively).

Let us consider an example:

In a study of the ability of rapid antigen tests to diagnose strep pharyngitis, 90% of patients with strep pharyngitis have a positive rapid antigen test, while only 5% of those without strep pharyngitis have a positive test. The LR+ for the ability of rapid antigen tests to diagnose strep pharyngitis is (select one):

$$\begin{aligned}\text{LR+} &= 90\% / 5\% = 18 \\ \text{LR+} &= 95\% / 10\% = 9.5 \\ \text{LR+} &= 90\% / 95\% = 0.95\end{aligned}$$

Interpreting likelihood ratios: general guidelines

The first thing to realize about LR's is that an LR > 1 indicates an increased probability that the target disorder is present, and an LR < 1 indicates a decreased probability that the target disorder is present. Correspondingly, an LR = 1 means that the test result does not change the probability of disease at all! The following are general guidelines, which must be correlated with the clinical scenario:

LR	Interpretation
> 10	Large and often conclusive increase in the likelihood of disease
5 - 10	Moderate increase in the likelihood of disease
2 - 5	Small increase in the likelihood of disease
1 - 2	Minimal increase in the likelihood of disease
1	No change in the likelihood of disease
0.5 - 1.0	Minimal decrease in the likelihood of disease
0.2 - 0.5	Small decrease in the likelihood of disease
0.1 - 0.2	Moderate decrease in the likelihood of disease
< 0.1	Large and often conclusive decrease in the likelihood of disease

The decision to order a test is also based on our initial assessment of the likelihood of the target disorder, and how important it is to rule-in or rule-out disease. For example, a chest x-ray might have a good likelihood ratio for pneumonia. But if you believe a patient has a simple cold, this test, no matter how good the LR, probably shouldn't be ordered. It is sometimes helpful to be able to calculate the exact probability of disease given a positive or negative test. We saw that this is next to impossible using sensitivity and specificity at the bedside (unless you can do Bayes' Theorem in your head). Next, we'll learn how we can use likelihood ratios to quickly estimate the probability of disease in our patients.

Relation between test performance characteristic

- Sensitivity and FNR

- Calculating sensitivity in terms of FNR
- Calculating of FNR in terms of sensitivity

- Specificity and FPR

- Calculating sensitivity in terms of FPR
- Calculating of FPR in terms of sensitivity

Summary of formula

True Pos = Sensitivity * Prevalence

False Neg = (1 - Sensitivity) * Prevalence

Pre Test Odds = Prevalence / (1 - Prevalence)

False Neg Rate = $100 * \text{False Neg} / (\text{True Pos} + \text{False Neg})$

4.11. Surveillance

Definition

Disease surveillance is the ongoing systematic collection, consolidation and analysis of data, and the dissemination of this information to those who need to know, so that actions may be taken. Surveillance is undertaken to inform disease prevention and control measures. Furthermore, it can be defined as surveillance is the continuous, systematic collection, analysis and interpretation of health-related data needed for the planning, implementation, and evaluation of public health practice.

"Ongoing, systematic collection, analysis, and interpretation of health-related data essential to the planning, implementation, and evaluation of public health practice, closely integrated with the timely dissemination of these data to those responsible for prevention and control."

Types,

1. Routine surveillance: In routine surveillance, the data are collected routinely and forwarded to higher levels on a routine basis. For this purpose, health staff of sub health post, health post and health center collect information about the number of cases of reportable communicable diseases including the EPI target diseases that occur in their health area. Health staff may also investigate each case of neonatal tetanus and polio. At the end of month they analyze the data and send it to the district supervisor.

2. Sentinel surveillance: It occurs when data are collected from only selected sites. This is rarely representative of population but can be used to monitor trends and collect more detailed information. Sentinel sites are chosen to report the number of cases of disease that occur for a specific time period and also in a certain age group. They may also provide a more consistent picture of illness in a given area. Data collected from these sites may also show whether the routine reporting data is working. In Nepal, Early Warning and Reporting and Response System has been developed for national priority diseases and different sentinel sites has been identified for the data collection.

3. Case/outbreak investigation: In general, case investigation is investigation of single case and an outbreak investigation is that of many cases. However, when the occurrence of a particular disease is very low, even one case can be considered an outbreak. The supervisors of concerned health facilities should investigate every case of neonatal tetanus and polio as soon as the case is suspected.

4. Special studies: These studies conducted by trained health professionals, investigators or epidemiologists. They are used to measure the number of cases of disease in an area and to evaluate the reliability of the routine or sentinel reporting.

Component and composition,

- Assessing public health status
 - To inform action in respect of the control and prevention of environmental hazards, exposures to potentially harmful agents, and the occurrence of disease

- Defining public health priorities
 - To inform policy and planning in respect of the current and likely future impact of hazards, exposures and disease
- Evaluating programmes
 - To inform decisions regarding existing interventions
- Stimulating research
 - To generate hypotheses and inform methodologies

Purpose

The surveillance is done for the following purposes:

- Observe disease trends and identify the disease as per case definition and causes
- Establish disease priority
- Identify, investigate and control outbreaks
- Identify, investigate and manage adverse events following immunization
- Document impact of services
- Estimate the requirement of vaccines/drugs and other logistics
- Evaluate quality of health service delivery
- Identify specific population groups at high risk of illness and deaths due to vaccine preventable diseases (VPDs) and other health problems
- Ensure that national immunization program (NIP) activities and resources are used efficiently
- Modify strategies for betterment

Importance

- Highlighting the magnitude of the illness as a public health problem
- Providing epidemiological data for planning program interventions
- Monitoring the quality of services
- Identifying high risk pockets for additional actions
- Identifying outbreaks early
- Estimating the requirement of vaccines for the programs
- Documenting impact of services to show

Process of surveillances

1. Collection of surveillance data
2. Analysis of surveillance data
3. Interpretation
4. Dissemination of surveillance findings

Sources of Data for Disease Surveillance

The following are the key sources of surveillance data:

- Mortality reports - Vital statistics, medical examiner data
- Morbidity reports - Notifiable disease reports, laboratory data, hospital data, outpatient health care data, specific disease data such as injuries data, adverse drugs reactions data, occupational illness data etc.

- Report of laboratory utilization (including laboratory test result)
- Report of individual case investigations
- Reports of epidemic investigations and epidemic
- Special surveys (e.g., hospital admission, disease registers and serologic surveys)
- Information on animal reservoirs and vectors
- Demographic data
- Environmental data

Surveillance system in Nepal

In Nepal, a system is in place for regular surveillance and reporting of suspected cases. Epidemiological services have been undertaken to collect the disease morbidity/mortality statistics from different disease endemic districts. There are two surveillance systems exist to report disease cases.

- Routine monthly reporting of aggregated data of all cases (suspected, probable, confirmed) from periphery to district, regional and central level (Health Management Information System, HMIS).
- Sentinel surveillance through Early Warning Reporting and Response System (EWARRS): Sentinel sites must include all clinical cases (suspected/probable/confirmed) in the immediate and weekly EWARS reporting forms, including "Zero" reports.

All outbreaks (suspected/probable cases) should be reported immediately to the respective regional health services directorate, the EDCC/DHS, and/or VBDRTC, Hetauda, for immediate investigation and, if possible, laboratory confirmation. The regional and central level should seek follow-up data on patient outcome. During outbreak situations, surveillance should be intensified with the introduction of active case finding whenever possible.

Laboratory confirmation should be performed as soon as possible. Thereafter, weekly reports of cases, age groups, deaths, regions, and hospital admissions should be set up. The surveillance has provided regular data for assessing the disease distribution, morbidity and mortality in different districts. Collected information has use for program implementation and policy implications. In Nepal, health staff from HPs/SHPs, PHCs and Hospitals collect information on the number of cases of reportable communicable diseases that occur in their catchment area in specific HMIS reporting forms. At the end of each month, they have to send this information to the DHO/DPHO, where this information is compiled by district and sent to the HMIS/Management Division, Department of Health Services

Early Warning Reporting and Response System in Nepal

Objective of EWARS

- to develop a comprehensive, computerized database of infectious diseases of public health importance
- to monitor and describe trends of infectious diseases through a sentinel

- surveillance network of hospitals followed by public health action and research
- to receive early warning signals of diseases under surveillance and to detect outbreaks
- to instigate a concerted approach to outbreak preparedness, investigation and response
- to disseminate data/information on infectious diseases through an appropriate feedback system.

4.12. Strategies of epidemiology

Ways and means of assembling facts on types of people affected by disease and by various circumstances

○ **Forming a hypothesis**

The descriptive epidemiology helps to formulate hypothesis relating to disease etiology. The epidemiological hypothesis should specify the population, the causes, the expected outcome, dose response relationship, time response relationship.

Descriptive characteristics are used to identify the possible factor(s) that caused the disease. The incidences or attack rates among various groups are compared. The risk group and the common source or sources to which the members were exposed are identified.

This step explains about the cause of the disease and suggests for blocking the progress of the disease, and serving as a guide for further investigation.

○ **Testing of hypothesis**

Analytical studies such as case control or cohort are carried out and data are analyzed using statistical test (e.g. attack rate, relative risk, or chi-square test) to determine probable sources. Differences between those exposed and unexposed to the risk factor of the disease is determined. Attributable risk is also calculated. These analysis help to search for other possible sources of infections or other modes of transmission. Epidemiologic findings should be confirmed by laboratory test of specimens such as blood, feces, suspected food etc.

Sometimes, various facts or hypotheses may fit with the result by analysis. Therefore, it is essential to test various hypotheses to ascertain whether it is consistent or inconsistent with all the known facts.

Standardization – rationale, direct and indirect standardization

„...is a set of techniques used to remove as far as possible the effects of differences in age or other confounding variables when comparing two or more variables.”

- The process by which you derive a summary figure to compare health outcomes of groups
- The process can be used for mortality, natality, or morbidity data

Methods of standardization

There are two methods of standardization commonly used in epidemiological studies, and these are characterized by whether the standard used is a population distribution (direct method) or a set of specific rates (indirect method). Both direct and indirect standardization involves the calculation of numbers of expected events (e.g. deaths), which are compared to the number of observed events.

1. Direct Method requires

- Age-specific rates in the sample population
 - The age of each case
 - The population-at-risk for each age group in the sample
- Age structure (percentage of cases in each age group) of a standard population
- Summary figure is an AGE-ADJUSTED RATE

2. Indirect method requires

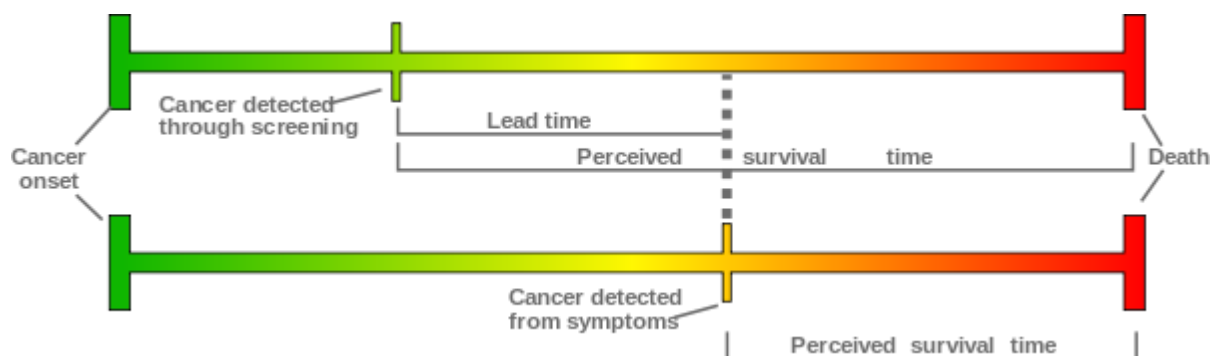
- Age structure of the sample population at risk
- Total cases in the sample population (not ages of cases)
- Age-specific rates for a standard population
- Summary figure is a STANDARDIZED MORTALITY RATIO (SMR)
- Instead of a standard population structure, you utilize a standard rate to adjust your sample
- Indirect standardization does not require that you know the stratum-specific rates of your cases
- The summary measure is the SMR or standardized mortality/morbidity ratio

$$\text{SMR} = \frac{\text{Observed}}{\text{Expected}} \times 100$$

An SMR of 100 or 100% means no difference between the number of outcomes in the sample population and that, which would be expected in the standard population

4.13. Lead time of diseases mechanism

Lead-time is the length of time between the detection of a disease (usually based on new, experimental criteria) and its usual clinical presentation and diagnosis (based on traditional criteria). OR It is the time between early diagnosis with screening, and when diagnosis would have been made without screening. Lead-time bias is that factor when comparing, in detection of disease, a traditional test with a new or experimental test but the outcome of the disease is unchanged; the new or experimental test merely identifies the disease earlier than the previous and thus gives the impression that survival is prolonged. It is an important factor when evaluating the effectiveness of a specific test.



4.14. Concept of risk, uses of risk, studies of risk and risk factors

“An exposure that is significantly associated with the development of disease” - WHO
 Risk factors are often suggestive, but absolute proof of cause and effect between a risk factor and disease is lacking, ie, the presence of risk factor doesn't sure that the disease will occur and in its absence the disease will not occur Eg:

- Smoking is risk factor for lung cancer
- Lack of physical exercise is a risk factor for coronary heart disease
- HTN, elevated cholesterol, alcohol is risk factor for stroke
- Alcohol, speed, road way design is risk factor for road traffic accident
- Obesity, diet is risk factor for Diabetes
- Alcohol is risk factor for cirrhosis of liver

Modifiable risk factor

- Smoking, alcohol, obesity, hypertension, physical activity, cholesterol level

Un modifiable risk factor

- Age, sex, genetic factors are cannot be changed

Risk Group

- The group which are more vulnerable to disease than others and need direct appropriate action to them and always in need of more attention

WHO guidelines for defining risk group for MCH services

- Biological situation
 - Age group: low birth weight, elderly primi
 - Short statured primi (140cm and below)
 - Malpresentation of fetus
- Disease condition during pregnancy
 - Anemic pregnant mother
 - APH, PPH
 - CVD, Kidney disease, diabetes, liver disease
 - Previous still birth

- Physical situation
 - Rural and urban slums
 - Poor living condition, overcrowding
 - Poor sanitation and water supply
- Socio cultural situation
 - Low social class people/ women
 - Uneducated women
 - poverty

4.14.1. Exposure

In a general sense, exposure can be defined as any of a subject's attributes (association) or any agent (effect) with which the subject may come into contact. These attributes or agents may be relevant to his or her health (Armstrong et al., 1998).

For an environmental factor, exposure can be more precisely defined as contact of some agent at the boundary between humans and the environment, at a specific concentration, over an interval of time (Wallace, 1995). Exposures can be harmful or beneficial.

Exposure Assessment:

the science that describes how an individual or population comes in contact with a risk factor, including quantification of the amount of the risk factor across space and time (Lioy, 1990).

- **Exposure intensity:** the agent/risk factor *concentration* in the medium that is in contact with the body
- **Exposure frequency:** designates *how often* the exposure occurs
- **Exposure duration:** the *length of the time* that the exposure occurs
- **Microenvironment:** defined as any *location or activity* in which a distinct exposure occurs.

When investigating whether an exposure is related to the risk of disease, the epidemiologist must consider many possibilities. Could the effect be related to the concentration? the frequency of exposure? duration? or is there something peculiar to a certain microenvironment?

Examples of exposure assessment

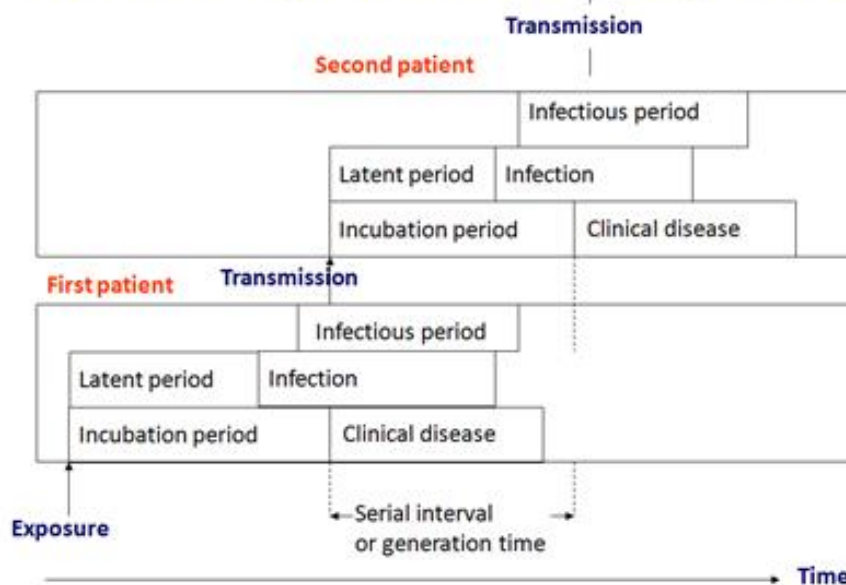
Agent(s)	biological, chemical, physical, single agent, multiple agents, mixtures
Source(s)	anthropogenic/non-anthropogenic, area/point, stationary/mobile, indoor/outdoor
Transport/carrier medium	air, water, soil, dust, food, product/item
Exposure pathway(s)	eating contaminated food, breathing contaminated workplace air touching residential surface

Exposure concentration	mg/kg (food), mg/litre (water), $\mu\text{g}/\text{m}^3$ (air), $\mu\text{g}/\text{cm}^2$ contaminated surface), % by weight, fibres/ m^3 (air)
Exposure route(s)	inhalation, dermal contact, ingestion, multiple routes
Exposure duration	seconds, minutes, hours, days, weeks, months, years, lifetime
Exposure frequency	continuous, intermittent, cyclic, random, rare
Exposure setting(s)	occupational/non-occupational, residential/non-residential, indoors/outdoors
Exposed population	general population, population subgroups, individuals
Geographic scope	site/source specific, local, regional, national, international, global
Time frame	past, present, future, trends

4.15. Latency period

Latent period: the period between exposure and the onset of infectiousness (this may be shorter or longer than the incubation period). The period between exposure and infection is called 'latent period', since the pathogen is present in a 'latent' stage, without clinical symptoms or signs of infection in the host. The period between exposure and onset of clinical symptoms is called 'incubation period'.

Relationships between time periods



4.16. Concept of prognosis of disease and prognostic factors

Definition

Prognosis is a prediction of the chance of recovery or survival from a disease. Most physicians give a prognosis based on statistics of how a disease acts in studies on the general population. Prognosis can vary with lung cancer depending on several factors, such as the stage of disease at diagnosis, type of cancer, and even gender.

Prognosis is a Statistic

Most information you will hear and read about the prognosis of your disease is based on statistics from studies looking at other people.

It's important to note that these numbers are only numbers, and do not look at individual variations. Most statistics are also somewhat dated. For example, statistics looking at the 5-year survival rate for a particular disease may be several years old—and since the time they were reported, newer and better treatments may have become available.

Prognosis is Different for Everyone

Every single cancer is different. If there are 200 people with stage 2A lung cancer in a room, there are 200 cancers that differ in molecular profiles and other important variants. On top of this, every person has important differences that affect prognosis, such as age, general health, co-existing medical conditions, and ability to tolerate treatment. Look at some of the many factors that can affect survival rate for people with lung cancer.

Factor that affects for Prognosis

1. **Survival rate** - The survival rate is the "average length of time someone is expected to survive a cancer, and is usually given based on a period of time, for example, "the 5-year survival rate."
2. **Median survival rate** - The median survival rate is a number which defines the time after which half of people with a certain type and stage of disease are alive, and 50% have died.
3. **Progression-free survival** - [Progression-free survival](#) is usually used to describe the response to a treatment for cancer, and refers to the average amount of time during which a cancer does not grow, or remains stable.

***Progression-free survival (PFS)** denotes the chances of staying free of disease progression for a group of individuals suffering from a cancer after a particular treatment. It is the percentage of individuals in the group whose disease is likely to remain stable (and not show signs of progression) after a specified duration of time. Progression-free survival rates are an indication of how effective a particular treatment is.*

Progression-free survival is often calculated for treatment of diseases that are slow growing and difficult to cure, like low grade lymphomas. This term is also used when salvage treatments are offered in situations where the intention is not cure but the control of disease.

4. **Disease-free survival** - [Disease-free](#) survival refers to the length of time that someone remains free of detectable disease.

Disease-free survival (DFS) denotes the chances of staying free of disease after a particular treatment for a group of individuals suffering from a cancer. It is the percentage of individuals in the group who are likely to be free of disease after a specified duration of time. Disease-free survival rates are an indication of how effective a particular treatment is.

Examples *'The 2-year disease-free survival for stage IIA Hodgkin lymphoma is 80% when treated with a new combination of drugs.' This means that after this particular treatment, about 80% of those treated are likely to be free of disease at 2 years.*

5. **Overall survival** - [Overall survival](#) refers to the average length of time someone survives after a diagnosis of cancer before death from any cause including cancer.

Overall survival, or OS, is a statistical term referring to the percentage of people in a group who are alive after a defined length of time – usually years. For instance, the 5-year overall survival rate may be reported for people with a particular cancer – in which case it is the percentage of people who are living 5 years after diagnosis or 5 years after the start of therapy.

Examples *"The 5-year overall survival for stage Leukemia is about 90 percent." This means that of all patients with Leukemia in the reported group, 90 percent lived at least 5 years after the cancer was diagnosed.*

These all prognosis factor are depends upon the treatment standard, diagnosis procedure, humal physiological status, immunological status of deseased population including their mental status and other social determinants parts.

4.17. Emerging and re-emerging infectious diseases and threat to clinical and public health

Definition

An emerging disease is one that has appeared in a population for the first time, or that may have existed previously but is rapidly increasing in incidence or geographic range.

■ Emerging infectious disease

Newly identified & previously unknown infectious agents that cause public health problems either locally or internationally

■ Re-emerging infectious disease

Infectious agents that have been known for some time, had fallen to such low levels that they were no longer considered public health problems & are now showing upward trends in incidence or prevalence worldwide

Factors Contributing To Emergence

AGENT

- Evolution of pathogenic infectious agents (microbial adaptation & change)
- Development of resistance to drugs
- Resistance of vectors to pesticides

HOST

- Human demographic change (inhabiting new areas)
- Human behaviour (sexual & drug use)
- Human susceptibility to infection (Immunosuppression)
- Poverty & social inequality
- Poor populations- major reservoir & source of continued transmission
- Poverty- Malnutrition- Severe infectious disease cycle
- Lack of funding, Poor prioritization of health funds, Misplaced in curative rather than preventive infrastructure, Failure to develop adequate health delivery system

ENVIRONMENT

- Climate & changing ecosystems
- Economic development & Land use (urbanization, deforestation)
- Technology & industry (food processing & handling)
- International travel & commerce
- Breakdown of public health measure (war, unrest, overcrowding)
- Deterioration in surveillance systems (lack of political will)
- Deforestation forces animals into closer human contact- increased possibility for agents to breach species barrier between animals & humans
- El Nino- Triggers natural disasters & related outbreaks of infectious diseases (Malaria, Cholera)
- Global warming- spread of Malaria, Dengue, Leishmaniasis, Filariasis

Population environment

- Growth of densely populated cities- substandard housing, unsafe water, poor sanitation, overcrowding, indoor air pollution (>10% preventable ill health)
- Problem of refugees & displaced persons
- Diarrhoeal & Intestinal parasitic diseases, ARI
- Lyme disease (*B. burgdorferi*)- Changes in ecology, increasing deer populations, suburban migration of population

Transmission of Infectious Agent from Animals to Humans

- >2/3rd emerging infections originate from animals- wild & domestic
- Emerging Influenza infections in Humans associated with Geese, Chickens & Pigs
- Animal displacement in search of food after deforestation/ climate change (Lassa fever)
- Humans themselves penetrate/ modify unpopulated regions- come closer to animal reservoirs/ vectors (Yellow fever, Malaria)

Human Behaviour

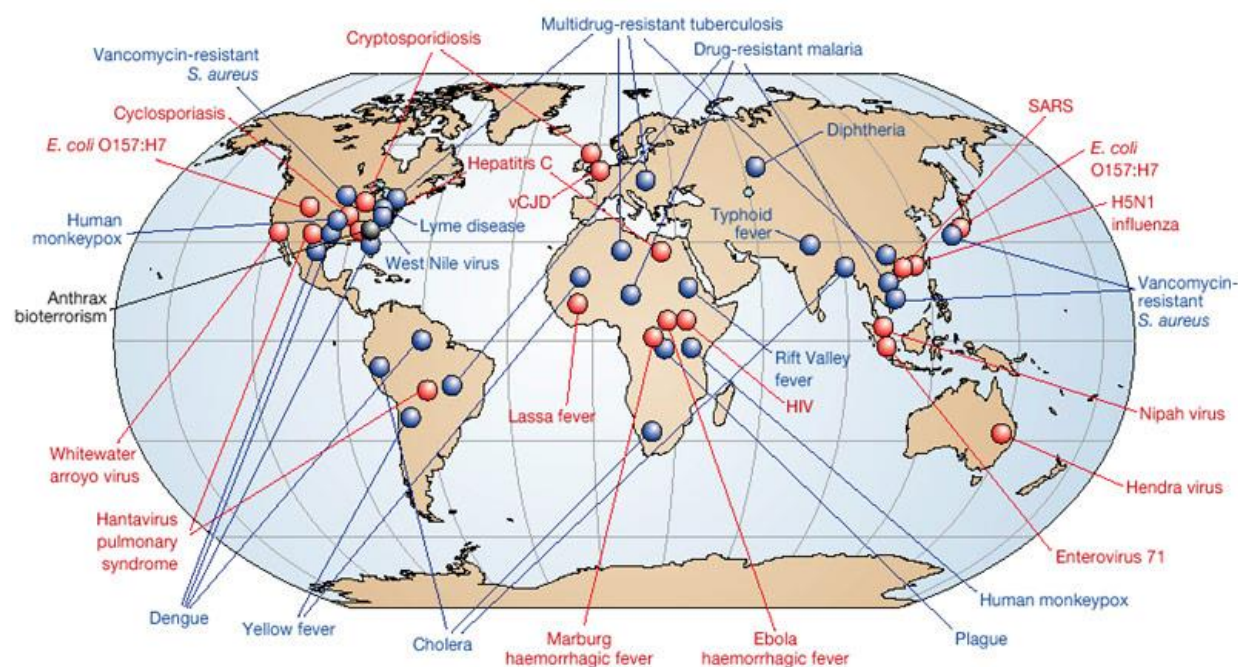
- Unsafe sexual practices (HIV, Gonorrhoea, Syphilis)
- Changes in agricultural & food production patterns- food-borne infectious agents (E. coli)
- Increased international travel (Influenza)
- Outdoor activity

Antimicrobial Drug Resistance

Causes:

- Wrong prescribing practices
- non-adherence by patients
- Counterfeit drugs
- Use of anti-infective drugs in animals & plants
- Loss of effectiveness:
- Community-acquired (TB, Pneumococcal) & Hospital-acquired (Enterococcal, Staphylococcal)
- Antiviral (HIV), Antiprotozoal (Malaria), Antifungal

Examples of recent emerging disease



CHAPTER 5

Genetic Epidemiology

Genetic Epidemiology

Genetic epidemiology is defined as the study of the distribution of and risk factors for diseases and genetic and environmental causes of familial resemblance. Genetic epidemiology focuses on how genetic factors and their interactions with other risk factors increase vulnerability to, or protection against, disease (Beaty 1997). Genetic epidemiology employs traditional epidemiologic study designs to explain aggregation in groups as closely related as twins or as loosely related as migrant cohorts.

Epidemiology has developed sophisticated designs and analytic methods for identifying disease risk factors. With increasing progress in gene identification, these methods have been extended to include both genetic and environmental factors (MacMahon and Trichopoulos 1996 ; Kuller 1979). In general, study designs in genetic epidemiology either control for genetic background while letting the environment vary (e.g., migrant studies, half siblings, separated twins) or control for the environment while allowing variance in the genetic background (e.g., siblings, twins, adoptees/ non biological siblings). Investigations in genetic epidemiology are typically based on a combination of study designs including family, twin, and adoption studies.

Application of Genetic epidemiology

- There is a widespread consensus among geneticists and epidemiologists on the importance of epidemiology to the future of genetics
- Best strategy for susceptibility risk factor identification for common and complex disorders will ultimately involve large epidemiologic studies from diverse populations
- It is likely that population-based association studies will assume increasing importance in translating the products of genomics to public health. There are several reasons that population-based studies are critical to current studies seeking to identify genes underlying complex disorders.
- Genes as risk factors is based nearly exclusively on clinical and nonsystematic samples. Hence, the significance of the susceptibility alleles that have been identified for cancer, heart disease, diabetes, and other common disorders is unknown in the population at large.
- Identification of risk profiles will require large samples to assess the significance of vulnerability genes with relatively low expected population frequencies.
- Its play role of epidemiology in quantifying risk associated with traditional disease risk factors, applications of human genome epidemiology can provide information on the specificity, sensitivity, and impact of genetic tests to inform science and individuals

Gene

Concept

A gene is a locus (or region) of DNA that encodes a functional RNA or protein product, and is the molecular unit of heredity. Glossary the transmission of genes to an organism's offspring is the basis of the inheritance of phenotypic traits. Most biological traits are under the influence of polygenes (many different genes) as well as the gene–environment interactions. Some genetic traits are instantly visible, such as eye colour or number of limbs, and some are not, such as blood type, risk for specific diseases, or the thousands of basic biochemical processes that comprise life.

Genes can acquire mutations in their sequence, leading to different variants, known as alleles, in the population. These alleles encode slightly different versions of a protein, which cause different phenotype traits. Colloquial usage of the term "having a gene" (e.g., "good genes," "hair colour gene") typically refers to having a different allele of the gene. Genes evolve due to natural selection or survival of the fittest of the alleles. The concept of a gene continues to be refined as new phenomena are discovered.

Classical concept of gene:

Classical concept of gene was introduced by sutton (1902) and was elaborated by morgan (1913). Bidge (1923), Muller (1927) and others which outlined as follows.

- a. Genes are discrete particles inherited in mendelian fashion that occupies a definite locus in the chromosome and responsible for expression of specific phenotypic character.
- b. Number of genes in each organism is more than the number of chromosomes; hence several genes are located on each chromosome.
- c. The genes are arranged in a single linear order like beads on a string.
- d. Each gene occupies specific position called locus.
- e. If the position of gene changes, character changes.
- f. Genes can be transmitted from parent to off springs.
- g. Genes may exist in several alternate formed called alleles.
- h. Genes are capable of combined together or can be replicated once during a cell division.
- i. Genes may under for sudden changes in position and composition called mutation.
- j. Genes are capable of self duplication producing their own exact copies.

Modern concept of gene:

S. Benzer (1957) coined different terms for different nature of gene and genetic material in relation to the chromosome on the basis of genetic phenomena to which they involve.

I. Genes as unit of transmission or cistron :

The part of DNA specifying a single polypeptide chain is termed as cistron. A cistron can have 100 nucleotide pairs in length to 30,000 nucleotide pairs. It transmits characters from one generation to other as unit of transmission.

II. Genes as unit of recombination or recon :

The smallest segment of DNA capable of being separated and exchange with other chromosome is called recon. A recon consists of not more than two pairs of unclesotides.

III. Gene as unit of mutation or muton :

Muton is the smallest unit of genetic material which when changed or mutated produce a phenotypic trait. Thus muton is delimited to a single nucleotide.

Gene Types

On the basis of their behavior the genes may be categorized into the following types.

- a. **Basic genes:** These are the fundamental genes that bring about expression of particular character.
- b. **Lethal genes:** These bring about the death their possessor.
- c. **Multiple gene:** When two or more pairs of independent genes act together to produce a single phenotypic trait.
- d. **Cumulative gene:** Some genes have additive effects on the action of other genes. These are called cumulative genes.
- e. **Pleiotropic genes:** The genes, which produce changes in more than one character, is called pleiotropic gene.
- f. **Modifying gene:** The gene which cannot produce a character by itself but interacts with other to produce a modified effect is called modifier gene.
- g. **Inhibitory gene:** The gene which suppresses or inhibits the expression of another gene is called inhibitory gene

Genetic disorder/disease

A genetic disease is any disease that is caused by an abnormality in an individual's genome, the person's entire genetic makeup. The abnormality can range from minuscule to major from a discrete mutation in a single base in the DNA of a single gene to a gross chromosome abnormality involving the addition or subtraction of an entire chromosome or set of chromosomes. Some genetic disorders are inherited from the parents, while other genetic diseases are caused by acquired changes or mutations in a preexisting gene or group of genes. Mutations can occur either randomly or due to some environmental exposure

There are a number of different types of genetic inheritance/disorder, including the following four modes:

1. Single gene inheritance

Single gene inheritance, also called Mendelian or monogenetic inheritance. This type of inheritance is caused by changes or mutations that occur in the DNA sequence of a single gene. There are more than 6,000 known single-gene disorders, which occur in about 1 out of every 200 births. These disorders are known as monogenetic disorders (disorders of a single gene).

Some examples of monogenetic disorders include:

- cystic fibrosis,
- sickle cell anemia,
- Marfan syndrome,
- Huntington's disease, and
- hemochromatosis.

Single-gene disorders are inherited in recognizable patterns: autosomal dominant, autosomal recessive, and X-linked

2. **Multifactorial inheritance**

Multifactorial inheritance, which is also called complex or polygenic inheritance. Multifactorial inheritance disorders are caused by a combination of environmental factors and mutations in multiple genes. For example, different genes that influence breast cancer susceptibility have been found on chromosomes 6, 11, 13, 14, 15, 17, and 22. Some common chronic diseases are multifactorial disorders. Examples of multifactorial inheritance include:

- heart disease,
- high blood pressure,
- Alzheimer's disease,
- arthritis,
- diabetes,
- cancer, and
- obesity.

Multifactorial inheritance also is associated with heritable traits such as fingerprint patterns, height, eye color, and skin color.

3. **Chromosome abnormalities**

Chromosomes, distinct structures made up of DNA and protein, are located in the nucleus of each cell. Because chromosomes are the carriers of the genetic material, abnormalities in chromosome number or structure can result in disease. Abnormalities in chromosomes typically occur due to a problem with cell division.

For example, Down syndrome (sometimes referred to as "Down's syndrome") or trisomy 21 is a common disorder that occurs when a person has three copies of chromosome 21. There are many other chromosome abnormalities including:

- Turner syndrome (45,X0),
- Klinefelter syndrome (47, XXY), and
- Cri du chat syndrome, or the "cry of the cat" syndrome (46, XX or XY, 5p-).

Diseases may also occur because of chromosomal translocation in which portions of two chromosomes are exchanged.

4. **Mitochondrial inheritance**

This type of genetic disorder is caused by mutations in the non-nuclear DNA of mitochondria. Mitochondria are small round or rod-like organelles that are involved in cellular respiration and found in the cytoplasm of plant and animal cells. Each mitochondrion may contain 5 to 10 circular pieces of DNA. Since egg cells, but not sperm cells, keep their mitochondria during fertilization, mitochondrial DNA is always inherited from the female parent.

Examples of mitochondrial disease include:

- an eye disease called Leber's hereditary optic atrophy;
- a type of epilepsy called MERRF which stands for myoclonic epilepsy with Ragged Red Fibers; and

- a form of dementia called MELAS for mitochondria lencephalopathy, lactic acidosis and stroke-like episodes.

How are genetic problems diagnosed?

Families at risk for genetic diseases may want to consult a certified genetic counselor. A careful family pedigree (chart of members of the family) and history may help determine risks for certain problems. Genetic counseling also helps parents understand the effects of a disorder and ways it may be prevented or treated.

It may be necessary to check each parent's DNA to learn about some genetic inheritance patterns. Prenatal testing is also available to check the fetus for problems. Testing may include ultrasound (using sound waves to look at internal structures), chorionic villus sampling (testing the tissues around the fetus), or amniocentesis (withdrawing a sample of the amniotic fluid).

What are teratogenic problems?

Certain substances are known to cause abnormalities in babies. Many birth defects occur when the fetus is exposed to teratogens (substances that cause abnormalities) during the first trimester of pregnancy when organs are forming. Some known teratogens include the following:

- Certain medications (always consult your doctor before taking any medications during pregnancy)
- Alcohol
- High level of radiation exposure
- Lead
- Certain infections (such as rubella)

CHAPTER 6

Social Epidemiology

Definition

Branch of epidemiology that studies the social distribution and social determinants of health (Berkman and Kawachi 2000). All epidemiology is social epidemiology (Kaufman and Cooper 1999) with the analysis of the social determinants of Health .

Builds and expands by posing new research questions, utilising new research methods and influencing government policy agenda.

History:

- 1674 John Graunt : who was dying during outbreaks .
- 1763 Louis René Villermé : social class & work condition as the crucial determinants of health & diseases.
- 1800 physician & others argued that bad air & emanation from decaying matter cause outbreak of illness , observing that deaths were commonly clustered among poor .
- 1830 John Snow who methodically covered the streets of London collecting statistics documenting the location of outbreak. Snow identified that contaminated water from communal pumps is the source of cholera
- History continued....
- 1844 Friedrich Engels : the horrible working condition & identified the relationship between these conditions & disease.
- 1860 Germ theory identifies single causal agent such as *Salmonella Typhi* as the cause of disease.
- By the 1950 -60 : clinician & epidemiologist monitoring these conditions to understand the chronic disease caused by combination of biological social behaviour patterns.
- 1960- the health impact of social condition & social class as the key determinants of morbidity & mortality .
- By the end of 20th century the concept of social causation of disease gained traction

DIFFERENCE BETWEEN MODERN & SOCIAL EPIDEMIOLOGY

MODERN EPIDEMIOLOGY

- Biological paradigm
- All diseases are biological phenomena can be described in fully biological paradigm.
- Disease is reflection of individual risk factor

SOCIAL EPIDEMIOLOGY

- Bio psychosocial paradigm
- Biology of organism is determined in multilevel interactive environment
- Diseases are assumed to be product of mutual interaction among social factor, individual factor & biological factor

Goals of Social epidemiology

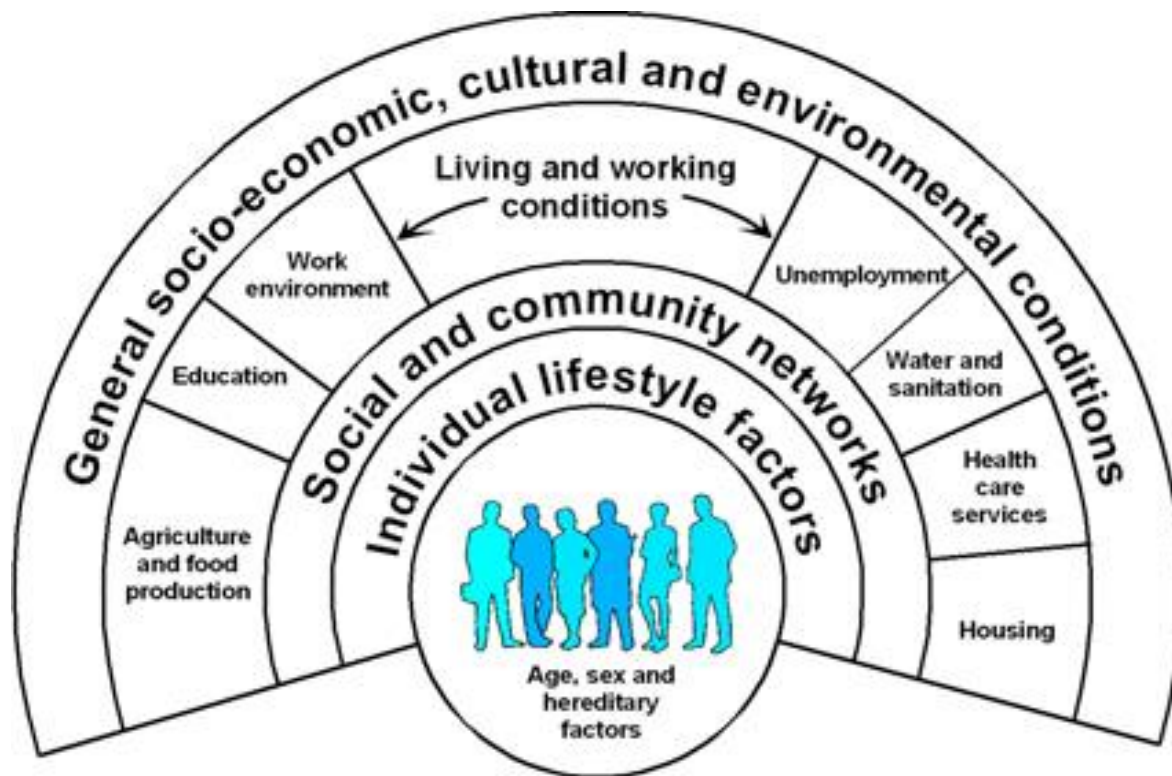
is to conceptualize, operationalize and test the associations between aspects of the social environment (families, workplaces, residential neighbourhoods, the political economy) and population health

- The range of problems studied by social epidemiologist : whether neighbourhood contexts , or workplace organization, or income inequality and social cohesion affect the health

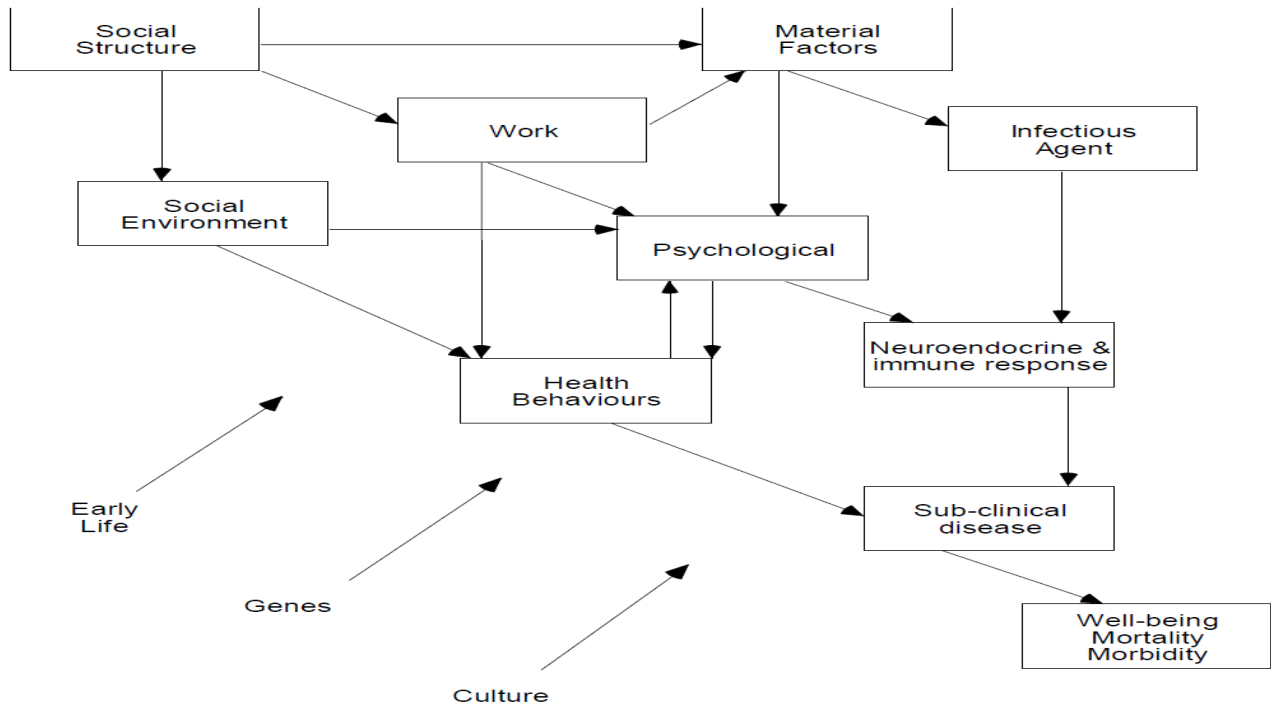
What are the social determinants of health?

"The poor health of the poor, the social gradient in health within countries, and the health inequities between countries are caused by the unequal distribution of power, income, goods, and services, globally and nationally, the consequent unfairness in the immediate, visible circumstances of peoples lives –their access to health care, schools, and education, their conditions of work and leisure, their homes, communities, towns, –and their chances of leading a flourishing life. This unequal distribution of health-damaging experiences is not in any sense a natural phenomenon....Together, the structural determinants and conditions of daily life constitute the **social determinants of health**."

(WHO Commission on Social Determinants of Health, 2008)

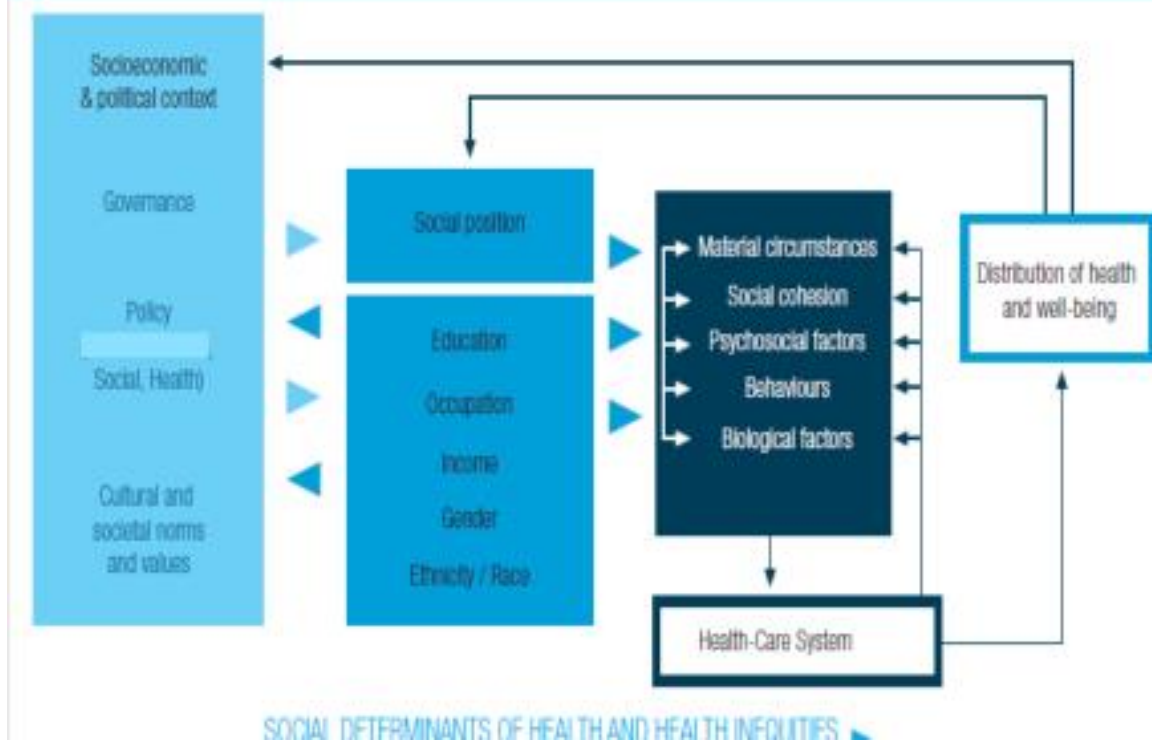


Pathway underlying Social determinant Marmot & Wilkinson (Below Figure)



Pathway underlying Social determinant Marmot & Wilkinson

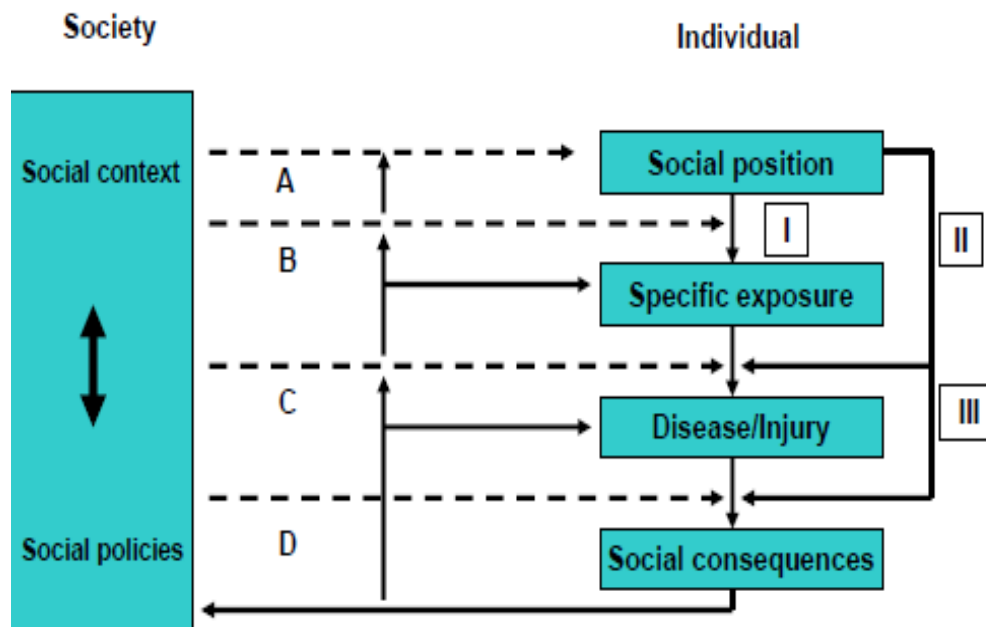
Figure 4.1 Commission on Social Determinants of Health conceptual framework.



Why emphasize social determinants?

- Have a direct impact on health.
- Predict the greatest proportion of health status inequity.
- Social determinants of health structure health behaviours.
- Interact with each other to produce health.

Upstream and downstream mechanisms of social inequities in health



Theories of Social Epidemiology in the 21st century Eco-social prospective

In social epidemiology the three main theoretical framework for explaining disease distribution are

1. Psychosocial
2. Social production of disease /political economy of health
3. Eco-social & other emerging multi level frameworks

Commission on social determinant of health (WHO) suggest how to closing the gap in generation

The commissioner overarching recommendation:

1. Improve daily living condition
2. Tackle the inequitable distribution of power money & resources
3. Measure & understand , assesses the impact of action

1. Improve daily living condition

Equity from start

- Commit & implement a comprehensive approach to early life building on existing child survival programmes & extending intervention in early life to include social emotional & language /cognitive development
- Expand the provision & scope of education to include the principal of early child development (physical social emotional & language).

Healthy places healthy people

- Place health & health equity at the heart of urban governance & planning
- Promote health equity between rural & urban areas through sustained investment in rural development
- Ensure that economic & social policy responses to climate change & other environmental degradation take into account health equity.

Fair employment decent work

- Make full & fair employment & decent work a central goal of national & international social & economic policy.
- Achieving health equity requires safe ,secure & fairly paid work year round work opportunities & healthy work balance for all
- Improve the working condition for all worker to reduce their exposure to material hazard , work related stress & health damaging behaviour

Social protection across life course

- Establish & strengthen universal comprehensive social protection policies that supports a level of Income sufficient for healthy living.
- Ensure social protection to those normally excluded .

Universal health care

- Build health care system based on principles of equity , disease prevention & health promotion
- Ensure that health care system financing is equitable .
- Build & strengthen the health workforce & expand capabilities to act on the social determinant of health

2. Tackle the inequitable distribution of power money & resources

Structural drivers of the conditions of daily life globally, nationally, and locally

Health equity in all policy system programmes

- Place responsibility for action on health & health equity at the highest level of governance & ensure its coherent consideration across all policies.
- Adopt a social determinant framework across the policy & programmatic function of the ministry of health & strengthen in supporting a social determinants approach across government.

Fair financing

- Strengthen public finance for action on the social determinants of health.

- Increase international finance for health equity & coordinate increased finance through social determinants of health action framework.
- Fairly allocate government resources for action on the social determinant of health

Gender equity

- Address gender biases in the structures of society –in laws & their enforcement in the way organisation are run & intervention designed.
- Develop finance policies & programme that closes gaps in education & skills & that support female economic participation.
- Increase investment in sexual & reproductive health services & programme building to universal coverage & right.

Political empowerment- inclusion of voice

- Empower all group of society through fair representation in decision making & how society operates particularly in relation to its effect on health equity .
- Enable civil society to organise & act in a manner that promotes & realize the political & social right affecting health equity

Good global governance

- Make health equity global development goal & adopt a social determinant of health framework to strengthen multilateral action for development.
- Strengthen WHO leadership in global action on the social determinant of health institutionalizing social determinants of health as a guideline principle across WHO department

3. Measure and Understand the Problem and Assess the Impact of Action

The Social Determinants of Health: Monitoring, Research and Training

- Ensure that routine monitoring systems for health equity and the social determinants of health are in place, locally, nationally, and internationally.
- Invest in generating and sharing new evidence on the ways in which social determinants influence population health and health equity and on the effectiveness of measures to reduce health inequities through action on social determinants.
- Provide training on the social determinants of health to policy actors, stakeholders, and practitioners and invest in raising public awareness

Examples of action

Sweden

- National health policy with a focus on decreasing health inequity based on population interventions defined with a social determinants approach .

Cuba

- Intersectoral approach to child health between health and education sectors resulting in strong interaction between health staff in polyclinics and other sectors, along with

emphasis on early child development with almost all children (99.8%) attending early child services.

- Cuba has very low child mortality across all groups and high educational attainment despite significant economic difficulties

New Zealand

- Whole-of-government national policy to reduce inequities led by health sector with primary health care reform, now showing reduction in major health inequity (between health status of indigenous and non-indigenous New Zealanders)

Thailand

- Implementation of universal health care coverage without fee-for-service, using a capitation based system with a primary health care approach

Brazil

- Implementation of Family Health Programme to improve coverage of health care using a health team approach, building in intersectoral action, which is already showing impressive improvements in infant mortality

Government of India

- Indira Aawas yojna
- Public distribution system
- Rajiv gandhi swasthya guarantee yojna
- Vridha Pension Yojna

Nepal

- Ama Surakshya Karayam Karam
- Garib ka lagi primeminister Karyakaram
- Old age, Widow and single women allowance etc
- Allowance for Kidney failure patients

Social Determinants of Health and Primary Health Care

- Advance holistic view of health, with primary value of health equity
- The Declaration of Alma implicitly referred to the social determinants
- Primary health care starts with the health sector and reaches out to other sectors
- Social determinants discourse sees health sector as one of the social determinants
- Report of the Commission and the upcoming World Health Report thus complement each other, and the Commission's findings will inform WHO's revitalisation of primary health care

How society affect health: area of research

- The social determinants of health:
- Health behaviours
- Material, economic and political determinants of health:
- Life course

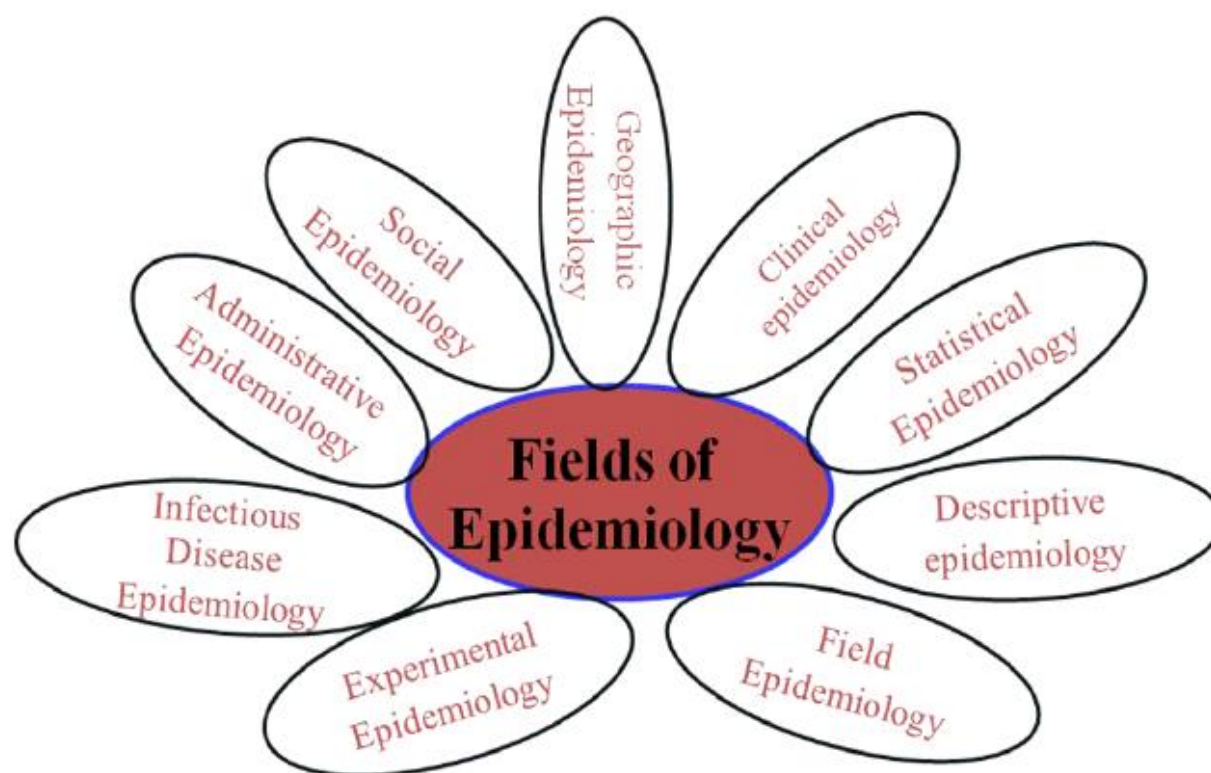
- Social biology
- Ecological prospective
- General susceptibility of disease
- Social support
- Social disorganisation
- Work stress
- Depression & affective disorder

CHAPTER 7

Field Epidemiology

7.1. Introduction

The history of epidemiology is far long which goes back to Hippocrates. The then, conceptual framework was developed to study the infectious disease and was limited to epidemic conditions only. The concept of epidemiology has been going on broadening with increasing its field. Principles of epidemiology are simply the guidelines for the epidemiology but its application may or may not rest on it. Thus the application of theories and principles of epidemiology in an actual field situation is known as field epidemiology. **Field epidemiology is the systematic process of measuring various field circumstances in which disease or other conditions of ill health tend to occur or not.** Field epidemiology is the collection of various activities such as outbreak investigation, data collection, surveying and displaying the results of investigation to those concerned people. It starts with the planning of activities for future which is known as **proposal development**. Thus field epidemiology is the practice of application of epidemiological methods and procedures in field setting for the control of existing problems and prevention emergence of health problems



7.1.1. Field Epidemiological Techniques

Field techniques include various procedures starting from the selection of area of study through the adequate mental analysis. Providing background and existing condition of health problems, stating the basic questions that one attempts to give answer (research questions), defining the objectives of study

through valid and scientific logic, searching for the similar evidences known as literature review, making the plan of study with anticipated results by using certain tools, construction of instruments of study, taking ethical clearance from relevant authorities etc.

- The application of theories and principles of epidemiology in an actual field situation
- Systematic process of measuring various field circumstances in which disease/condition tends to occur/not occur. Eg, Epidemic/Outbreak Investigation, including observational and interventional studies
- Applied form of epidemiological theories and principles
- Practice/Application of epidemiological methods to control and prevent health problems

7.1.2. Field Epidemiology: General Concept

- Field investigation is action oriented, with a prime goal to solve a pressing public health problem
- Epidemiologists are mobilized under a variety of circumstances
 - Primarily when
 - A problem is acute and unexpected
 - Quick action is required
 - Problem is pressing; Eg, involvement of press, political pressure
- Primary thought: The need to institute the controls necessary to safeguard health

7.1.3. Challenges for Field Investigation

- Limited control over the situation
- Little time for planning a study
- Limited information for solving the problem-Eg, data sources and laboratory samples

7.2. Field Techniques (Steps of proposal development)

- Include various procedures starting from
 - Problem identification
 - Developing research question
 - Statement of problems
 - Literature review
 - Study objectives
 - Study design
 - Construction of test instrument
 - Ethical clearance ...

7.2.1. Field techniques include various procedures such as:

- The selection of area of study through the adequate mental analysis.
- Providing background and existing condition of health problems, stating the basic questions that one attempts to give answer (research questions),
- Defining the objectives of study through valid and scientific logic,
- Searching for the similar evidences known as literature review,
- Making the plan of study with anticipated results by using certain tools,
- Construction of instruments of study,
- taking ethical clearance from relevant authorities

1. Problem Identification

The first and foremost step in field epidemiology is to identify the existing problem. There are various health problems in the community; of them selection and refining the study topic is of paramount importance. It provides the area of study and background of problem. Problems are the difficulties which the epidemiologist experiences in the context of either theoretical or practical situation and wants to obtain the solution for the same. The problems should be identified through adequate exercise and understandings. Thus, a problem is one which requires a researcher to find out the best solution for the given problem.

2. Proposal Development The development of proposal or working guide or protocol is important. It is the first step in research work which guides for action. A proposal is a written document for the purpose of obtaining fund from granting agencies or approval from instructional review boards which contains the study protocol and their administrative and supportive information that required by the specific agencies. It is the working guideline for actions to achieve specific results. Writing and submitting a research proposal is the challenging act for researcher. It is the inquisitiveness for all researchers. This is frustrating and discouraging for new and learning researchers whereas a step of progress for experienced researcher. It aids in knowledge and analytical power of researcher.

A proposal document consists of following essential components

1. *Research problem*
2. *knowledge of relevant facts*
3. *Methodology*
4. *ethical clearance*
5. *Anticipated results and their implication*
6. *Administrative and management guide*
7. *Study team*

Guideline for development of proposal

- For the successful outcomes and better utilization of available resources to prepare appropriate proposal, one should follow the prescribed guideline. This is suggestive but not exhaustive i.e. the steps may differ depending upon the nature of problem and style of work. Basic points remember are as follows

First step of epidemiological study/field epidemiology

1. Organize a team of relevant expert and divide responsibility
2. Prepare the proposal on prescribed format according to the requirement of concerned institute.
 - a written document
 - for the purpose of
 - obtaining fund from granting agencies or
 - approval from instructional review boards (NHRC in Nepal)
3. Develop the protocol for priority problems and should give Forces investigator to
 - Organize,
 - clarify and

- refine all the elements of the study
- 4. State the topic, objectives, and methodology in clear and simple terms
 - Simple, SMART based and sound methodology should be choosed for field investigation
- 5. Select optimum sample through appropriate technique.
 - Ensure that sampling frame and size are adequate to reach the valid conclusions
- 6. Review, pretest and revise frequently.
- 7. Take ethical clearance from related individuals and institutions
- 8. Work according to time table.
 - Establish a timetable and meet periodically
- 9. Apply creative and sound report writing and presentation skills

More Advances in Proposal Development According to Condition arise

- Development of Test Instruments
- Tools used to obtain information required to fulfill the research/study objectives
 - Questionnaire
 - Observation checklist
 - Interview schedule
- Development usually occurs before but sometimes after approval of the proposal depending upon the policies of the funding organization
 - Only name of test instruments/tools mentioned in the latter case
- Careful construction of test instruments is necessary to avoid missing of any study variables
- Pretest the tool(s) among suitable respondents
- Appropriate modification may be needed

Steps of proposal development Introduction

- Background
- Statement of problem
- Rationale of Study
- Significance of Study
- Objectives (General and Specific)
- Research Questions
- Hypothesis of study
- Conceptual framework
- Operational Definition

2. Review of Literature

- Theoretical review
- Research review

3. Methodology of Study

- Type of study/study design
- Study area
- Study population
- Sampling technique

- Sample size
- Tools and techniques of information collection
- Data analysis, management and Interpretation
- Validity and Reliability of the Research
- Supervision and Monitoring plan
- Inclusion and exclusion criteria
- Expected outcome
- Ethical consideration
- Time table for work

A short explanation of some important step of proposal development is stated below

1. Introduction of study

Background of study

It provides the basic information regarding the research topic and problem. It is a means of brief description of the topic. It should start with historical background or broader aspects of topic and then specific contents. It states the existing situation of health problem which is supported by facts and figures. Background of study is simple description about the problem of major title, the area where research going to be applied, and about the target population. It depicts just about the epidemiological trend of problem, socio demographic and socioeconomic and socioculture condition of study area also.

Statement of Problem

It provides the logical and evidence based information about the causation of arising as research study problem in study area.

- The first step in the condition and matter of development of a research project
- To state the research problem precisely and clearly
- To clarify and focus and evidence based information
- An integral part of selecting a research topic

Rationale of study

The principles or reasons on which formulation of research is based. Rationale of study proves for research and its importance. Rationale or justification is describing the causes of doing research. Rationale always plunk on the evidence based cause. Based on rationale the researcher derives the importance/ significance of study and objective of study.

Significance of study

Simply, significance of study is importance of study that implies future meaning of result and its application. Significance of study always coincide with outcome or output of study that connote the significantly verification of rationale and objective. Prove of significance of study make a research viable.

Objective study

Objectives are end results that the researcher wants to achieve. Objectives must be stated in terms of General and specific types. The valid justification of topic and its use must be specified in this portion. Objective of study should be SMART. 'What information are we going to collect in our study to meet our objectives?'

S	Specific
M	Measurable
A	Achievable
R	Reasonable
T	Time bound

Research objectives are the goals to be achieved by the application of a research project

• Two types:

1. **General Objective:** What is to be accomplished by the research project and why?
2. **Specific Objective(s):** in detail, the specific aims of the research project, often breaking down what is to be accomplished into smaller logical components. Specific objectives relate to the specific research questions the investigator wants to answer through a proposed study.

Just for Example

a. General objectives

To determine whether there is a causal association between a vasectomy and subsequent hospitalization due to atherosclerotic diseases, and, other predisposing risk factors for coronary disease.

b. Specific objectives

- *To estimate the overall relative risk of vasectomy as well as other risk factors for atherosclerotic diseases in men*
- *To estimate the independent effect of vasectomy on atherosclerosis (using a conditional logistic regression model)*
- *to test the possible duration of the effect of vasectomy on risk for atherosclerosis and*
- *To examine the possible synergistic effect between vasectomy, cigarette smoking and hypertension*

Research Question

Research question defined the clear idea about the findings and research objective. It gives the instruction for study design. It depicts about the aim of research. Characteristics of research question Specific, Clear, Definite, depict the whole research/generalization of research

Hypothesis

Research Hypothesis:

- A tentative prediction or explanation of the relationship between two or more variables
- Prediction of expected outcomes
- It translates the problem statement into a precise, unambiguous prediction of expected outcomes
 - Hypothesis should reflect the depth of knowledge, imagination and experience of the investigator and are not meant to be haphazard guesses.
 - A hypothesis can be as simple in form as predicting the relationship between two variables, one independent and one dependent. Therefore, in the process

of formulating hypotheses, all variables relevant to the study should be identified.

Examples: "Health education based on active participation of mothers produces more positive changes in child feeding than health education based on lectures" Independent variable: types of health education Dependent variable: change in child feeding

Conceptual framework

Conceptual framework is overall arrangement of total research; basically it tends to describe the relation ship between dependent and independent variable. It provides clearly about the study variable, matter and mode or scale of measurement of relationship of cause and effect.

- Variable: A concept which can take on different quantitative or qualitative values [Eg, weight, height, income; smoking habit, lung cancer, employment status, sex etc.]
- In conceptual framework basically we used Four major types of variables
 1. **Independent variables:** Variables that are manipulated or treated in a study in order to see what effect differences in them will have on those variables proposed as being dependent on them. Independent variables are **antecedent** to the dependent variable. [Syn: cause, input, risk factor, determinant and attribute]
 2. **Dependent variables:** Variables in which changes are results of the level or amount of the independent variable(s). Dependent variables are **consequences** of the other (independent) variables [Syn: effect, outcome, consequence, result, condition and result]
 3. **Confounding/intervening variables:** Variables that may influence or confound the effect of independent variable on dependent variable [Eg, in the study of the effect of measles (IV) on child mortality (DV), the nutritional status of the child may play an intervening role]
 4. **Background variables:** Variables those are so often relevant in investigations of population that they should be considered for possible inclusion in the study [Eg, sex, age, ethnic origin, education, marital status, social status, etc...]

Clarification of Independent and Dependent Variable In research, particularly analytical and interventional study (research hypothesis testing study), the term independent and dependent variable are used.

The independent variable is referred to as the experimental treatment. The researcher controls what treatment will be selected and how much will be applied. The treatment, or independent, variable will not change during the research or as a result of the research. The dependent variable is the one that is expected to change as a result of the treatment. It is not under the control of the researcher. Said another way, the independent variable is expected to cause some effect on the dependent variable. The changed, or effected, variable is referred to as dependent because its value depends on the value of the independent variable. Actually, the independent variable forms or defines groups; the dependent variable generates data.

2. Review of literature It incorporates the review of already known facts and correlation with new facts and is compared with research as a source of study is literature review. .

Types of literature review

1. Based upon the publication
 - Published Article
 - Unpublished Article
2. Based upon the of study
 - Theoretical review
 - Research review
3. Based upon the research
 - Review article
 - Original article

Purpose of literature review:

- To get familiarized with existing knowledge about the research problem and
- To find out if others have investigated the same/similar problems

Importance of review of literature

1. It helps further understanding of the problem proposed for research and may lead to refining of –the statement of the problem;||
2. it helps to identify the study variables and conceptualize their relationships;
3. it helps in the formulation and selection of research hypotheses;
4. it helps in finding out what others have reported on the topic so that account may be taken of this design of the research;
5. it provides with familiarity with the various methods that might be used in the research

Source of information for literature review

- Card catalogues of books in libraries;
- Indices, such as the Index Medicus and the International Nursing Index which identify journal articles by subject, author and title;
- Computer-based literature searches such as MEDLINE, MEDLARS and CATLINES;
- Bibliographies, such as those found at the end of books, articles and these, or prepared as separate documents;
- Statistics collected at national, provincial and/or departmental levels and
- Responses to enquiries about ongoing research
- A thorough search of body of literature in scientific databases (such as PubMed and Cochrane library) and via personal communication with the experts
- Followed by a critical review of the collected literature

3. Methodology of study

Methodology of study is the procedural guideline to reach at desired results. It includes various steps such as selecting appropriate study design, choosing study area, calculating desired sample size, describing procedures of study and selecting appropriate methods and

tools of study. The development of study instrument and its pre-testing is of paramount importance for the collection of accurate information.

Study or Research design: Research design is very important part of research body. It deals with the strategy of research process and specific study methodology to achieve exact objective. The whole research depends upon the study design wheatear the design is suitable for the research title and objective or not. It includes,

- 1) Selection of research strategies
- 2) It is the core of research design, and is probably the single most important decision the investigator has to make.
- 3) The choice of strategy, whether descriptive, analytic, experimental, operational or combination of these, depends on a number of considerations.

The specific types of studies are as follows:

- Descriptive strategies (observational hypothesis generation rather than testing)
- Observational analytical strategies (hypothesis testing)
- Experimental strategies
- Operational strategies

The major research design using,

- Descriptive cross-sectional study or population survey, e.g. malaria survey, opinion survey, knowledge-attitude-practice (KAP) survey;
- Epidemiological description of disease occurrence by person, place and time;
- Studies of changing patterns of health and disease over time and space: the epidemiological translation;
- Community diagnosis of a health problem or assessment of needs;
- Studies of existing data: case-series, disease registries, surveillance reports;

Descriptive study

- Case study and case series
- Cross sectional descriptive

Observational analytical strategies

- Prospective or retrospective study (cohort study)
- Historical (or reconstructed) cohort study, when adequate historical data or records are available;
- Retrospective study (case-control study);
- Analytical cross-sectional study;
- Follow-up study (longitudinal study; repeated cross-sectional study).

Experimental strategies

- Animal studies;
- Therapeutic clinical trials,
- Prophylactic clinical trials;
- Field trials;
- Community trial

- Quasi
- RCT
- Quasi-experimental studies (intervention studies, health systems research)

Study Area

Area, where, we are going to implement the research program.

Study Population

Target populations where, we going to implement the research. Study populations are eligible as criteria of research. From the study population researcher selected the sample for conducting research. So, study population is generalization group of sample of research program.

Sample size

Sample is group that represents the whole study population. The study design based upon the sample size. There are different methods to estimate the sample for study. Sample is depends upon the strength of study, burden of study, epidemiological status of study, study design, total population etc. It is different for experimental, descriptive and observational study.

Sampling

- 1) The process/technique of selecting a sample of appropriate and manageable size for study.
- 2) The process of obtaining information about an entire population by examining only a part of it.
- 3) The selection of some part of an aggregate or totality on the basis of which a judgment or inference about the aggregate or totality is made

Sampling include

- Selection of probability sampling method - simple random; systematic and stratified sampling; cluster sampling; multiphasic; multistage, sequential; repetitive; weighted and stratified
- Determination of sample size – the sample should be of sufficient size to produce meaningful results and to allow tests of statistical significance to be applied.
- The plans should be made to ensure representativeness and reliability of the sample to minimize sampling errors

Tools and technique of Information collection Study instruments - Tools by which data are collected in the study:

- a) Questionnaire and interview schedules
- b) preparation and pre-testing of questionnaires;
- c) plan for interviews and call backs;
- d) preparation of instruction manual;
- e) Training of interviewers.
- f) Other methods of observation
- g) medical examination;

- h) laboratory tests;
- i) screening process)

4. Development of study Instruments

Development of research instrument requires a good analytical mind. Frame the test instruments by keeping in mind that whether the objectives are addressed or not. These include questionnaire, observation check list, interview schedule focuses group discussion guideline etc. These are the tools by which information can be collected to fulfill the research objectives. These should be constructed carefully and appropriate modification may be needed after its pretest among the similar population. Sometime the construction of research instrument is the former step before the submission to concerned department and refining is done after its review. Definition and redefinition of instruments is essential for the valid research results.

Methods of data collection: This depends largely upon what type of data is being collected for the study. There are two major types of data:

1. Primary data: are those which are collected afresh and for the first time, and thus happen to be original in character.
2. Secondary data: are those which have already been collected by someone else and which have already been passed through the statistical process.

There are several methods of collecting primary data. Important ones are

- A. Observation method
- B. Interview method
- C. Through Questionnaires
- D. Through schedule

A. Observation method: The information is sought by way of investigator's own direct observation without asking from the respondent.

Advantages:

- a. The main advantage of this method is that subjective bias is eliminated, if observation is done accurately.
- b. Secondly, the information obtained under this method relates to what is currently happening; it is not complicated by either the past behavior or future intentions or attitudes.
- c. Thirdly, this method is independent of respondents' willingness to respond and as such is relatively less demanding of active cooperation on the part of respondents as happens to be the case in the interview or the questionnaire method.
- d. This method is particularly suitable in studies which deal with subjects (i.e. respondents) who are not capable of giving verbal reports of their feelings for one reason or the other.

Limitations:

- a. It is an expensive method.
- b. The information provided by this method is very limited.

- c. Sometimes, unforeseen factors may interfere with the observational task.
- d. At times, the fact that some people are rarely accessible to direct observation creates obstacle for this method to collect data effectively.

Types of observation:

I. **Participant observation/Non-participant or disguised observation:** distinction depends upon the observer's sharing or not sharing the life of the group he is observing. If the observer observes by making himself, more or less, a member of the group he is observing so that he can experience what the members of the group experience, the observation is called as the participant observation. But when the observer observes as a detached emissary without any attempt on his part to experience through participation what others feel, the observation of this type is often termed as non-participant observation. When the observer is observing in such a manner that his presence may be unknown to the people he is observing, such an observation is described as disguised observation.

Merits of participant type of observation:

- The researcher is enabled to record the natural behavior of the group.
- The researcher can even gather information which could not easily be obtained if he observes in a disinterested fashion.
- The researcher can even verify the truth of statements made by informants in the context of a questionnaire or a schedule.

Demerits of participant observation:

- The observer may lose the objectivity to the extent he participates emotionally,
- The problem of observation-control is not solved and
- It may narrow-down the researcher's range of experience

II. **Controlled/Uncontrolled observation:** if the observation takes place in the natural setting, it may be termed as uncontrolled observation, but when observation takes place according to definite pre-arranged plans, involving experimental procedure, the same is then termed controlled observation.

B. Interview Method

Schedule Questionnaire

I. Definition

II. Types of questionnaire:

According to structure

- a. Structured
- b. Unstructured
- c. Semi structure

According to types of questions

- d. Open-ended
- e. Close-ended
 - i. Dichotomous

- ii. Multiple choice question
- iii. Rating scale
- iv. Numerical
- v. Listing

According to mode of administration

- a. Mailed
- b. Telephone
- c. Face-to-face
- d. Self

Advantages and disadvantages of each types.

III. Formatting of questionnaire

- 1) Language and wording style
 - i. Leading question
 - ii. Offensive question
 - iii. Forced choice question
 - iv. Ambiguous question
- 2) Coding responses to questions
- 3) Cascade format of question
- 4) Length of questionnaire
- 5) Reliability of questionnaire
- 6) Validity of questionnaire
- 7) Layout of questionnaire
- 8) Sequencing of questions in questionnaire

IV. Steps in questionnaire administration:

- 1) Pretesting or pilot testing
- 2) Training interviewers
- 3) Call back
- 4) Editing and coding

V. Qualities of good questionnaire

VI. What types of information can be collected by using questionnaires?

Introduction to questionnaire

A questionnaire is simply a list of mimeographed or printed questions that is completed for the respondents. The use of questionnaires is a standard method of data collection in epidemiological research. Questionnaires are the back bone of the epidemiological research so a good questionnaire is the basis for the conduction of research in the field. Questionnaires must fulfill the aims, the correct format, appropriate language and wording style. The length of the questionnaires should be reasonable, it should be reliable and valid.

Types of questionnaires

- Mailed questionnaires
- Telephone based questionnaires
- Face to face

- Standard questionnaires or inventories questionnaires
- Factors to be considered for questionnaire design
- Study objectives & major research questions
- Study hypotheses: what data are required to accept or reject a hypothesis
- Data to be collected
- Plans for analysis and dummy tables so that no important information is missed
- Budget and
- The audience or target population: age, sex, stranger in house, can a wife be interviewed in the absence of the husband?) Above all will the respondents be able to give the required answers?
- Questionnaire Format

- 1) Structured questionnaires: The questionnaires that consists of close ended or structured questions.
- 2) Unstructured questionnaires: In this type of questionnaires the open ended or unstructured questions are used to obtain the information needed.
 - Forms of structured questions

Structured questions may offer

- a) A dichotomous choice of _yes' or _no', 'approve' or _disapprove, or _effective or _not effective'. Questions of this type should always include a _don't know' response category.
- b) Multiple choice of items
- c) A rating scale
- d) A numerical answer e.g. _How old are you?

Qualities of Questionnaire Following are some characteristics of good questionnaire:

- It should be of appropriate length.
- It should be in logic order.
- Simple, clear and concise language should be used.
- Relevant questions should be used as per hypothesis and objectives
- Easier question should be asked first.
- Avoid embarrassing questions or it should be parsed in appropriate language or synonym.
- Suggestive questions should be avoided.

Analysis, data management and interpretation

Analysis of data:

Data collection is followed by **data analysis**. It involves establishment of categories, the application of these categories to raw data through coding, tabulation and then drawing statistical inferences. Researcher should classify the raw data into some purposeful and usable categories.

Coding is the process of transforming the categories of data into symbols that may be tabulated or counted.

Editing is the procedure that improves the quality of the data for coding. Tabulation follows coding process.

Tabulation is a part of the technical procedure wherein the classified data are put in the form of tables. The mechanical devices can be made use of at this juncture. Computers can be used to deal with a great deal of data especially in large studies. Computers not only save time but also make it possible to study large number of variables affecting a problem simultaneously.

Tabulation is generally followed by computation of various percentages, coefficients, etc. by applying various well defined **statistical formulae**. In the process of analysis, relationships/differences supporting/conflicting with original or new hypotheses should be subjected to tests of significance to determine with what validity data can be said to indicate any conclusion(s). For example, if there are two samples of weekly wages, each sample being drawn from factories in different parts of the same city, giving two different mean values, then our problem may be *whether the two mean values are significantly different or the difference is just a matter of chance*. Through the use of statistical tests, we can establish whether such a difference is a **real one** or is the result of **random fluctuations**. If the difference happens to be real, the inference will be that the two samples come from different universe and if the difference is due to chance, the conclusion would be that the two samples belong to the same universe. In conclusion, various statistical measures help in analysis of the collected data.

Generalizations and interpretation:

Generalization: Making inference(s) to the population from study findings derived from the study on sample derived from the same population. If a hypothesis is tested and upheld several times, it may be possible for the researcher to arrive at generalization, i.e., to build a theory. Interpretation: *Synonyms: analysis, understanding, explanation*. It may be defined as a process of explaining the findings on the basis of some theory. It may trigger off new questions which in turn may lead to further researches.

- To observe how the language used in the questions is understood by different groups in the population
- In developing health care interventions, such as health education campaigns, FGD can be used to see how messages are received, understood and communicated to others.

Disadvantages

1. Groups may be inhibiting for some participants, who may feel unable to give their opinions, particularly if they are likely to be marginal to those of the rest of the group.
2. They can take considerable time and effort to set up.
3. Need for skilled moderators
4. Recruitment of participants can be slow
5. Analysis of collected data can be difficult and time consuming

References

1. AB Joshi, Fundamentals of epidemiology vol I and II.
2. Beaglehole R, Bonita R, Kjellstrom T. Basic Epidemiology. World Health Organization, Geneva, 2002.
3. Gordis L. Epidemiology, Second Edition, WB Saunders Company, Aharcourt Health Sciences Company, Philadelphia, 2000.
4. Park K. A text book of preventive and social medicine; Latest edition, Jabalpur: Banarasidas Bhanot Publishers.
5. Abraham M Lilianfield, David E Lilienfield "Foundation of Epidemiology", Oxford University Press.
6. Heninkens CH, Buring JE. Epidemiology in Medicine, Lippincott Williams and wilkins, a Wolters Kluwer Company: 1987.
7. MacMahon B, Trichopoulos D. Epidemiology: Principles and Methods, Second Edition. Boston: Little, Brown, 1996.
8. Mahajan BK. A Text Book of Preventive and Social Medicine
9. Principles of Epidemiology, 2nd Edition, and An Introduction to applied Epidemiology and Biostatistics. US Department of Health and Human Service, CDC, Atlanta Georgia.
10. Rothman KJ, Greenland S. Modern Epidemiology, 2nd Edition, Lippincott- Raven publishers: 1998.
11. Rothman KJ. Epidemiology: An Introduction. Oxford University Press, 2002

Principles and Practices of
Fundamentals and Clinical Epidemiology
(Reference Lecture Note)